GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

M nucleic - protein search, using frame_plus_n2p model

un on: October 9, 2003, 12:12:11 ; Search time 78 Seconds
(without alignments)
2779.028 Million cell updates/sec

itle: US-09-518-842-1
erfect score: 774
equence: 1 ATGGCCAGCCGTTCCGGTC.........CAGTTAAATTATGTACATAG 420

coring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop   6.0 , Delext 7.0

earched: 830525 seqs, 258052604 residues

otal number of hits satisfying chosen parameters: 1661050

inimum DB seq length: 0
aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

ommand line parameters:
MODEL=frame+_n2p.model -DEV=xlp
Q=/cgn2_1/USPTO_spool/US09518842/runat_09102003_111036_13065/app_query.fasta_1.583
DB=SPTREMBL_23 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
JSER=US09518842 @CGN_1_1_172 @runat_09102003_111036_13065 -NCPU=6 -ICPU=3
NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

atabase : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

esult No.	Score	Query Match	Length	DB	ID	Description
1	154.5	20.0	199	6	Q9N0T9	Q9not9 camelus dro

No.	c	Score	Match	Length	DB	ID	AC	Description
2		145	18.7	180	6	Q9MYX8	Q9MYX8	Q9myx8 felis silve
3		103	13.3	188	5	Q9GK46	Q9GK46	Q9gk46 galago cras
4	c	101	13.0	575	5	Q22328	Q22328	Q22328 caenorhabdi
5	c	100.5	13.0	146	11	Q8BT12	Q8BT12	Q8bt12 mus musculu
6		94	12.6	754	16	Q8ZQC3	Q8ZQC3	Q8zqc3 salmonella
7		94	12.6	754	16	Q8Z802	Q8Z802	Q8z802 salmonella
8		88.5	11.4	411	4	Q8TC33	Q8TC33	Q8tc33 homo sapien
9		88.5	11.4	754	2	Q9NUP6	Q9NUP6	Q9nup6 homo sapien
10		87.5	11.3	1218	11	Q9RLW4	Q9RLM4	Q9rlw4 klebsiella
11		85	11.0	685	11	Q8CIS1	Q8CIS1	Q8cis1 mus musculu
12		84	10.9	612	11	Q8CID1	Q8CID1	Q8cid1 mus musculu
13		83	10.7	321	16	Q9HZB9	Q9HZB9	Q9hzb9 pseudomonas
14	c	82.5	11.1	718	6	Q9BE45	Q9BE45	Q9be45 bos taurus
15		82.5	10.7	1626	5	O64293	O64293	O64293 streptococc
16	c	82	11.0	335	9	Q25340	Q25340	Q25340 leishmania
17		82	10.6	672	11	Q8CG10	Q8CG10	Q8cg10 mus musculu
18		81.5	10.6	795	11	Q9DBY0	Q9DBY0	Q9dby0 mus musculu
19		81.5	10.5	303	11	Q8K0P9	Q8KOP9	Q8k0p9 mus musculu
20		81.5	10.5	476	5	Q961W9	Q961W9	Q961w9 drosophila
21		81.5	10.5	604	11	Q8CDN3	Q8CDN3	Q8cdn3 mus musculu
22		81	10.5	1554	9	O36168	O36168	O36168 streptococc
23		81	10.5	803	16	Q8P500	Q8P500	Q8p500 xanthomonas
24		80.5	10.4	457	16	Q8ZB58	Q8ZB58	Q8zb58 yersinia pe
25		80.5	10.4	462	10	Q81712	Q8L7L2	Q81712 arabidopsis
26		80.5	10.4	463	16	Q8D1R0	Q8D1R0	Q8d1r0 yersinia pe
27		80.5	10.4	889	16	Q9FHH9	Q9FHH9	Q9fhh9 arabidopsis
28		80	10.3	111	4	Q9H2B5	Q9H2B5	Q9h2b5 homo sapien
29		79.5	10.3	183	10	Q8RUD8	Q8RUD8	Q8rud8 oryza sativ
30		78	10.1	311	11	Q8CBI0	Q8CBI0	Q8cbi0 mus musculu
31		78	10.1	1345	11	Q8CIP5	Q8CIP5	Q8cip5 mus musculu
32		77.5	10.0	192	6	Q8MKEI	Q8MKEI	Q8mkei macropus eu
33	c	77.5	10.4	490	16	Q8EYX2	Q8EYX2	Q8eyx2 leptospira
34	c	77.5	10.4	1217	5	P91457	P91457	P91457 caenorhabdi
35		77	9.9	136	11	Q9D718	Q9D718	Q9d718 mus musculu
36	c	77	10.3	577	10	Q9ATG6	Q9ATG6	Q9atg6 polytomella
37		77	9.9	836	6	O19057	O19057	O19057 pongo pygma
38		77	9.9	1417	12	Q89241	Q89241	Q89241 western equ
39		77	9.9	1490	12	Q88791	Q88791	Q88791 western equ
40	c	76.5	10.3	372	4	Q8NDB1	Q8NDB1	Q8ndb1 homo sapien
41	c	76.5	10.3	402	10	Q82202	Q82202	Q82202 arabidopsis
42	c	76.5	10.3	801	10	Q8IV53	Q8IV53	Q8iv53 homo sapien
43		76	9.8	188	11	O70148	O70148	O70148 rattus norv
44		76	9.8	740	4	Q8NC88	Q8NC88	Q8nc88 homo sapien
45		76	9.8	821	4	Q96JM3	Q96JM3	Q96jm3 homo sapien

ALIGNMENTS

RESULT 1
Q9N0T9
ID   Q9N0T9;   PRELIMINARY;   PRT;   199 AA.
AC   Q9N0T9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Preprorelaxin.
OS   Camelus dromedarius (Dromedary) (Arabian camel).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.
OX   NCBI_TaxID=9838;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20191540; PubMed=10727251;
RA   Hombach-Klonisch S., Abd-Elnaeim M., Skidmore J.A., Leiser R.,
RA   Fischer B., Klonisch T.;
RT   "Ruminant relaxin in the pregnant one-humped camel.";
RL   Biol. Reprod. 62:839-846(2000).
CC   -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR   EMBL; AF253739; AAF67741.1; -.
DR   InterPro; IPR004825; Ins/IGF/relax.
DR   Pfam; PF00049; Insulin; 1.
DR   SMART; SM00078; IlGF; 1.

DR   PROSITE; PS00262; INSULIN; 1.
SQ   SEQUENCE   199 AA; 22386 MW; F69BBD7E4F4A8089 CRC64;

Alignment Scores:
Pred. No.:   9.72e-09   Length: 199
Score: 154.50   Matches: 54
Percent Similarity: 34.17%   Conservative: 14
Best Local Similarity: 27.14%   Mismatches: 70
Query Match: 19.96%   Indels: 61
B: Gaps: 5

US-09-518-842-1 (1-420) x Q9N0T9 (1-199)

Qy   1 ATGGCCAGCCGTGTTCCGGTCCTATCTGCCAGCAATCTGGCTGCTGAGCCAACTCCTT 60
          ||| ||| |||||||| ::|||||||||||||||||||||||
Db   1 MetProArgLeuLeuLeuSerHisLeuLeuGlyValTrpLeuLeuLeuSerGlnLeuPro 20

Qy   61 AGAGAAAGCCTAGCAGCAGAG--------CTGAGGGGATGTGGTCCCCGATTTGGA 108
          :::|:|||| :::|: |||||
Db   21 LysGluThrSerGlyGluArgSerAsnAspPheValLysAlaCysPheValArgGluLeuVal 40

Qy   109 AAACACTTGCTGTCATATTGC--------CCCATGCCTGAGAAGACA 147
          ::: |||| ::: |||
Db   41 ArgLeuTrpIleGluIleCysGlySerValSerTrpGlyArgProAlaProArgProAla 60

Qy   148 TTCACCACCACCCCAGGAGGTGGCTG-------- 174
          |||
Db   61 ProArgProAlaProLysProAlaLeuArgProAlaLeuSerGlnAspLysLysProArg 80

Qy   175 CTGGAATCTGGACGTCCCAAAGAAATGGTGTCAACCTCCAACACAAAGATGGACAAGCC 234
          ||| |||||| ||||:::|| ||||||
Db   81 LeuArgSerGlyProProAlaGluIleMetProSerSerIleThrLysAspAlaGluThr 100

Qy   235 TTAGTACGACACATCAGAATTCATTCCTAATTGTCACCAGAGCTGAAGAAACCACTGTCT 294
          ||| ||| |||||| ||||||
Db   101 LeuThrThrMetLeuGluPheThrProAsnLeuProGlnLeuThrAlaThrLeuSer 120

Qy   295 GAAGGGCAGCCATCA-------- 309
          :::
Db   121 GluArgGlnProSerAlaGluProGlnGlnProAlaLeuLysAspSerAsnLeuAsnPhe 140

Qy   310 ----TTGAAGAAAATAATACTTCCCG-------- 333
          |||||||||
Db   141 GluGluPheLysLysIleIlePheAspArgGlnAsnGluGluAspGluSerGluSerLeuSer 160

Qy   334 ----AAAAGAGAAGTGACGTCACAGA 357
          ||||||||
Db   161 GluLeuLysAsnLeuGlyLeuAspLysHisSerGluLysGlyArgGlnLeuGlnMetThr 180

Qy   358 TTTGATCCATTCTGTTGTGAAGTAATTTGTGACCATGAAACTTCAGTAAAATATGT 414
          ||||||::: |||||
Db   181 LeuGlyGluArgCysCysGlnLysGlyCysSerSerArgLysGluMetAlaThrAlaCys 199

RESULT 2
Q9MYK8
DR   Q9MYK8   PRELIMINARY;   PRT;   180 AA.
AC   Q9MYK8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Relaxin.
OS   Felis silvestris catus (Cat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX   NCBI_TaxID=9685;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99115216; PubMed=9915995;
RA   Klonisch T., Hombach-Klonisch S., Froehlich C., Kauffold J.,
RA   Steger K., Huppertz B., Fischer B.;
RT   "Nucleic acid sequence of feline preprorelaxin and its localization
RT   within the feline placenta.";
RL   Biol. Reprod. 60:305-311(1999).

RN   [2]
RP   SEQUENCE FROM N.A.
RA   Hombach-Klonisch S., Klonisch T.;
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR   EMBL; AF233688; AAF60303.1; -.
DR   HSSP; P04090; 6RLX.
DR   InterPro; IPR004825; Ins/IGF/relax.
DR   Pfam; PF00049; Insulin; 1.
DR   SMART; SM00078; IlGF; 1.
DR   PROSITE; PS00262; INSULIN; 1.
SQ   SEQUENCE   180 AA; 20360 MW; 4C2CF371C698AF9F CRC64;

Alignment Scores:
Pred. No.:   1.17e-07   Length: 180
Score: 145.00   Matches: 53
Percent Similarity: 36.13%   Conservative: 16
Best Local Similarity: 27.75%   Mismatches: 58
Query Match: 18.73%   Indels: 64
DB: Gaps: 5

US-09-518-842-1 (1-420) x Q9MYK8 (1-180)

Qy   1 ATGGCCAGCCGTGTTCCGGTCCTATCTGCCAGCAATCTGGCTGCTGAGCCAACTCCTT 60
          ||| ||| |||||||| :::|||||||| ::||||||||||||
Db   1 MetLeuArgLeuPheLeuSerHisLeuLeuGlyValTrpLeuLeuLeuSerLeuArgAla 20

Qy   61 AGAGAAAGCCTAGCAGCAGAG--------CTGAGGGGATGTGGTCCCCGATTTGGAAAACAC 114
          |||:::| |||| ||||| |||||::: ||||
Db   21 ArgLysIleProAlaGlnAspGluGluValLeuLeuAlaCysGlyArgGluPheValArgLeu 40

Qy   115 TTGCTGTCATATTGCCCCCATG--------CCT 138
          ::: |||
Db   41 GlnIleArgIleCysGlySerLeuSerTrpGlyLysSerSerGlnHisArgGluPro 60

Qy   139 GAGAAGACATTCACCACCACCCCAGGAGGTGGCTGCTGGAATCTGGACGTCCCAAAGAA 198
          ::: |||
Db   61 ArgGlnAlaProAlaAlaLeuPro 69

Qy   199 ATGGTGTCAACCTCCAACACAACAAAGATGGACAAGCCTTAGTAGGTACGACACATCAGAATTCATT 258
          :::|| ||| |||| ::||| ||:::
Db   70 IleValSerSerSerIleThrSerGlyAlaAlaGluAlaLeuAsnGlyMetLeuGluTyrIle 89

Qy   259 CCTAATTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCA-------- 309
          |||:::||| ||||||||| ||||
Db   90 ProAspLeuProGlnGluLeuLysAlaThrLeuSerGluArgGlnProSerPheArgGlu 109

Qy   310 ----------------------------TTGAAGAAAATAATACTT 327
          ||||||
Db   110 LeuGlnProSerLeuLysAspSerAsnLeuAsnLeuGluGluValLysSerIleLeuLeu 129

Qy   327 --------------------------- 327
          |||
Db   130 GlyArgGlnAsnGluAlaGluAspGlnSerLeuSerGlnLeuGlyLysArgSerArgLeuAsp 149

Qy   328 -------TCCCGCAAAAGAGAAGAGTGACGTCACAGATTTGATCCATTCCTTGTGTGAAGTA 381
          |||::: |||||||||||| ||::::
Db   150 AlaHisSerArgIleLysArgSerArgTyrIleArgTyrSerAspArgCysCysAsnVal 169

Qy   382 ATTTGTGACGATGGAACTTCAGTTAAAATATGT 414
          ||| |||||
Db   170 GlyCysThrArgLysGluLeuLeuAlaAspLeuCys 180

RESULT 3
Q9GK46
ID   Q9GK46   PRELIMINARY;   PRT;   188 AA.
AC   Q9GK46;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Prorelaxin.
OS   Galago crassicaudatus (Thick-tailed galago) (Otolemur crassicaudatus).

C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C Mammalia; Eutheria; Primates; Strepsirrhini; Galagonidae; Otolemur.
X NCBI_TaxID=9463;
N [1]
P SEQUENCE FROM N.A.
C TISSUE=Placenta;
A Klonisch T., Froehlich C., Tetens F., Fischer B., Hombach-Klonisch S.;
T "Molecular remodeling of members of the relaxin family during primate
T evolution.";
L Mol. Biol. Evol. 0:0-0(2001).
C -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
C -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
R EMBL; AF317625; AAG42218.1; -.
R HSSP; P04090; 6RLX.
R InterPro; IPR004825; Ins/IGF/relax.
R Pfam; PF00049; Insulin; 1.
R SMART; SM00078; IlGF; 1.
R PROSITE; PS00262; INSULIN; 1.
Q SEQUENCE   188 AA;   21573 MW;   77551629B82B9A66 CRC64;

lignment Scores:
red. No.:   0.00772
core: 103.00
ercent Similarity: 34.50%
est Local Similarity: 24.00%
uery Match: 13.31%
B: 6

Length:	188			
Matches:	48			
Conservative:	21			
Mismatches:	57			
Indels:	74			
Gaps:	9			

S-09-518-842-1 (1-420) x Q9GK46 (1-188)

Y   1 ATGGCCAGCCTGTTCCGGTCCTATCTGCCAGCAATCTGGCTGCTGCTGAGCCAACTCCTT   60
       ||| ||||| |||||| |||:::
b   1 MetProArgLeuPhePheHisLeuLeuGlyValTrpLeuLeuLeuThrGlnIleSer   20

y   61 AGA------------------GAAAGCCTAGCAGCAGAGCTGAGGGGATGGTGGTCCCCGA 102
        ||| ||||| |||||| ||| |||
b   21 ArgAlaLysMetAspLysGlyGluAsnLeuAsnGlnIleIlePheAlaCysGlyArgArg   40

Y 103 TTTGGAAAACACTTGCTGTCATATTGCCCCATGCCTGAGAAGACAATTCACCACCACCCCA 162
        ||||||||| ||||
b   41 LeuIleArgIleTrpValGluValCys---------------------------------   49

Y 163 GGAGGGTGGCGTGCTGGAATCTGGACGTCCCAAA-------------------------- 195
        |||||| ||||| ||
b   50 ---GlySerThrGlyPheArgGlyAlaLysAsnGlnIleIleIlePheAlaCysGlyArg   68

Y 196 -----------GAAATGGTGTCAACCTCC---AACAACAAAGATGACAAGACCCTTAGGT 240
              || |||:: |||
b   69 GluProPheSerGluIleValProSerPheIleAsnLysAspAlaGluThrIleAsn   88

Y 241 ACGACATCAGAGAATTCATTCCTAATTG----------------------------- 267
       ||||||||| ||| ||||
b   89 MetMetSerGluPheIleAlaAsnLeuProGlnLysGlnThrGlnSerGluMet   108

Y 268 --------------TCACCAGAGCTGAAGAAA------------CCACTGTCTGAAGGG 300
                     ::::::||| :::|||
b 109 AsnLeuProSerProGluGlnIleTyrProThrLeuLysGlySerAspIleSer   128

Y 301 --CAG 303
                                                               |||
b 129 PheGluGluValLysAsnAsnIleHisAsnGluGlnGlyGluAlaGluAspAsnSerHis 148

Y 304 CCATCATGAAGAGAAAATATATACTTCCCGCAAAAAGAGAAGTGGACGTCACAGATTT--- 360
        ||||::: ||| ::: |||
b 149 SerGluLeuGlnAsnLeuGlyLeuAspThrHisSerArgLysLysSerArgGluArgTyrMet 168

Y 361 GATCCATTC-------TGTTGTGAAGTAATTTGTGACGATGAAACTTCAGTTAAATTATGT 414
        ||| ::: |||
b 169 SerProLeuGlnLysCysCysArgIleGlyCysThrLysArgSerLeuAlaArgPheCys 188

ESULT 4
22328
D Q22328   PRELIMINARY;   PRT;   575 AA.

AC Q22328;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein T07H6.5.
GN T07H6.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA Waterston R.;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Geisel C.;
RT "The sequence of C. elegans cosmid T07H6.";
RL Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; U53344; AAA96225.2; -.
DR HSSP; P10998; 1VVD.
DR WormPep; T07H6.5; CE31364.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 8.
DR SMART; SM00032; CCP; 8.
KW Hypothetical protein.
SQ SEQUENCE   575 AA;   63326 MW;   B6820F6C692AA646 CRC64;

Alignment Scores:
Pred. No.:   0.0171
Score: 101.00
Percent Similarity: 38.56%
Best Local Similarity: 28.10%
Query Match: 13.05%
DB: 5

Length:	575			
Matches:	43			
Conservative:	16			
Mismatches:	38			
Indels:	56			
Gaps:	9			

US-09-518-842-1 (1-420) x Q22328 (1-575)

Qy   94 GGTCCCGATTTGGAAAACACTTGCTGTCATATTGCCCATGCCTGAGAAGACATTCACC 153
        ||||||| ||| |||||||||| ||||:::
Db 429 GlyProProMet---LysCysLeuProSerTrpCysGluHisProSerLysThrTyrGly 447

Qy 154 ACCACCCCAGGAGGGTGGCGTGCTG------------------------------ 177
        ||| ||| ||||||||||||||||
Db 448 ThrLeuProGlyGlyGlnIleLeuLeuGluGlyLyIleLeuGlyAlaTyrGluPheGlnSer 467

Qy 178 -----------------------GAATCTGGACGTCCCAAAGAATGTGTCAACCTCCAACAACAAA 222
                              ||||| |||
Db 468 TyrIleGlnLysValGluGluGlyArg-------AlaIleSerPheGlnIleCysGlyLys 484

Qy 223 GATGACAAGACCCTTAGGT-----------------------ACGACATCAGAATTCATTCCTAAT 264
        ||| |||| :::|||
Db 485 GlyAsnTyrLeuIleGlyProProLysAlaThrCysValAsnGlyLysGluTrpMetProLys 504

Qy 265 TTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCATTGAAG--AAAATA 321
        :::||| ||| ||| ||| :::
Db 505 ValSerPro-------LysCysValSerGlnThrHisProMetIleGluGlyLysLysIle 521

Qy 322 ATACTTCCCGCAAAAAGAGAAGT-------GGACGT 351
        ::: |||||
Db 522 LeuTrpAspArgLysArgSerLeuProGlyArgAlaValArgGluTyrValAspAsp 541

Qy 352 -------------------------CACAGATTTGATCCATTCTGTTGTGAAGTAATT------- 384
                                 :::

GN   YCAI OR STM0983.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX   MEDLINE=21534948; PubMed=11677609;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT   LT2.";
RL   Nature 413:852-856(2001).
DR   EMBL; AE008742; AAL19917.1; -.
DR   InterPro; IPR001279; Blactmase-like.
DR   InterPro; IPR004477; ComEC_N-term.
DR   InterPro; IPR004797; ComEC_Rec2.
DR   Pfam; PF03772; Competence; 1.
DR   Pfam; PF00753; lactamase_B; 1.
DR   TIGRFAMs; TIGR00360; ComEC_N-term; 1.
DR   TIGRFAMs; TIGR00361; ComEC_Rec2; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   754 AA;   84954 MW;   711A95D282271358 CRC64;

Alignment Scores:
Pred. No.:   0.116   Length: 754
Score: 94.00   Matches: 42
Percent Similarity: 39.13%   Conservative: 21
Best Local Similarity: 26.09%   Mismatches: 54
Query Match: 12.62%   Indels: 45
DB: Gaps: 9

US-09-518-842-1 (1-420) x Q8ZQC3 (1-754)

Qy   399 AGTTCCATCGTCACAAATTACTTCACAACAGAATGGATCAAATCTGTGACGTCCACTTCT 340
         ::::::::: ||| |||||| ||| |||||| |||||||| |||||
Db   386 ThrAlaMetLeuAlaAsnLeuPheAlaValProTrpValThrProValThrProLeu 405

Qy   339 CTTTTTGCGGGAAAGTAT------------------------TATTTTCTTCAATGATGGCTGCCC 298
         ::: ||||| :::|||||| ||::
Db   406 IleLeuAlaGlyMetIleLeuHisLeuThrGlyProPheLeuGluTrpVal--- 424

Qy   297 TTCAGACAGTGGTTCTT------------------------------280
         :::||||| |||
Db   425 --------------TrpTyrLeuThrAspAlaArgAlaLeuAlaAlaAlaSerLeuAla 441

Qy   279 -----CAGCTTCGGTGACAAATTAGGAATGAATAGTCTG----ATGTCGTACCTAAGGCTT 230
              ||| ||| ||||||| |||||| :::::| |||
Db   442 LeuProGlnGlyTrpValAsnIleAspAsnArgTrpGlnTrpLeuThrLeuLeuProTrp 461

Qy   229 GTCCATCTTGTGTTGTTGGAGGTTGACACCATTCTTCTTGGGACGTCCAGATTCCAGCAGCC 170
         ::: ||| |||||||| ::||| --ProAlaVal
Db   462 LeuThrLeuIleAlaTrpAsn------ValTrpArgThrTrp--ProAlaVal 478

Qy   169 ACCCTCCTGGGGTGGTGGTGAATGTCTTCAGGCATGGGCAATATGACAGCAAGTGTT 110
         ||| ||| ||||||||||| ||||||
Db   479 CysLeuCysGlyLeuLeuLeuMetSerTrpProLeuTrpArgProIleAsnAlaSerGly 498

Qy   109 TTCCAAATCGGGAGCCACCACCATCCCCTCAGCTCGTCCTGCGTCTGCTAGGGGTTGGC 50
         :::||||:: |||::::
Db   499 TrpGlnVal-----HisMetLeuAlaAspVal------------------------GlyGlnGly 509

Qy   49 TCAGCAGCCAGCCAGATTGCTCGCAGATAGG------------------ACCGGAACAGGCTGG 5
         ||| ||| |||| :: ||:::
Db   510 LeuAlaIleAlaIleAlaArgGlyAlaSerLysPlysValIleLeuArgTyrAspThrGlyAlaArgAlaTrp 529

Qy   4   CCA 2
         |||
Db   530 Pro 530

RESULT 5
Q8BT12
ID   Q8BT12   PRELIMINARY;   PRT;   146 AA.
AC   Q8BT12;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Relaxin.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Body;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK028199; BAC25808.1; -.
SQ   SEQUENCE   146 AA;   16203 MW;   0E6B8AE2BDC6B08A CRC64;

Alignment Scores:
Pred. No.:   0.0141   Length: 146
Score: 100.50   Matches: 32
Percent Similarity: 36.28%   Conservative: 9
Best Local Similarity: 28.32%   Mismatches: 31
Query Match: 12.93%   Indels: 41
DB: Gaps: 2

US-09-518-842-1 (1-420) x Q8BT12 (1-146)

Qy   196 GAAATGGTGTCAACCTCCACAACAACAAAGATGGACAAGCCTTAGGTACCACACATCAGAATTC 255
         ||||:::||| ::||||||||| |||||
Db   32 GluValValProSerPheAsnLysAspAlaGluProPheAspThrThrLeuLysCys 51

Qy   256 ATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGAGGCAGCCATCATTGAAG 315
         :::|||||||||| ||||||| |||||||||
Db   52 LeuProAsnLeuSerGluGluLysAlaValAlaValAlaSerGlnAlaSerLeuPro 71

Qy   316 AAAATA--------ATACT--------------------------------------327
         |||
Db   72 GluLeuGlnHisAlaProValLeuSerAspSerValValSerLeuGluGlyPheLysLys 91

Qy   327 --327

Qy   92 ThrLeuHisAspLysLeuGlyGluAlaGluAspGlySerProProGlyLeuLysTyrLeu 111
Db   112 GlnSerAspThrHisSerArgLysArgGluSerGlyGlyLeuMetSerGlnGlnCys 131

Qy   373 TGTGAAGTAATTTGTGACGATGGAACTTCAGTTAAATTA 411
         ||| :: |||
Db   132 CysHisValGlyCysLeuSerArgArgSerIleAlaLysLeu 144

RESULT 6
Q82QC3
ID   Q82QC3   PRELIMINARY;   PRT;   754 AA.
AC   Q82QC3;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Putative recombination protein.

Db   542 GluLeuSerThrHisArgGlnHisSerGlyLysCysGlyIleValSerGlyLysGluGlu 561
         ||||| :::::
Qy   385 ------------------------TGTGACGATGGAACTTCAGTT 405
                                 |||::::||| |||||
Db   562 ArgMetIleMetGlnHisSerAspAsnGlyValSerVal 574

RESULT 7

Q8Z802
ID   Q8Z802   PRELIMINARY;   PRT;   754 AA.
AC   Q8Z802;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Putative competence-related protein.
OS   Salmonella typhi.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Salmonella.
OX   NCBI_TaxID=601;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CT18;
RX   MEDLINE=21534947; PubMed=11677608;
RA   Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA   Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA   Baker S., Basham D., Davies K., Chillingworth T., Connerton P.,
RA   Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA   Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA   Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA   Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA   Whitehead S., Barrell B.G.;
RT   "Complete genome sequence of a multiple drug resistant Salmonella
RT   enterica serovar Typhi CT18.";
RL   Nature 413:848-852(2001).
DR   EMBL; AL627268; CAD05383.1; -.
DR   InterPro; IPR001279; Blactmase-like.
DR   InterPro; IPR004477; ComEC N-term.
DR   InterPro; IPR004797; ComEC_Rec2.
KW   Hypothetical protein; Complete proteome.
  - Pfam; PF03772; Competence_1.
RN   Pfam; PF00753; lactamase_B; 1.
RN   TIGRFAMs; TIGR00360; ComEC N-term; 1.
RN   TIGRFAMs; TIGR00361; ComEC_Rec2; 1.
SQ   SEQUENCE 754 AA; 84920 MW; 6697C033FB29576E CRC64;

Alignment Scores:
Pred. No.:   0.116   Length: 754
Score: 94.00   Matches: 42
Percent Similarity: 39.13%   Conservative: 21
Best Local Similarity: 26.09%   Mismatches: 54
Query Match: 12.62%   Indels: 45
DB: 16   Gaps: 9

US-09-518-842-1 (1-420) x Q8Z802 (1-754)

Qy   399 AGTTCCATCGTCACAAATTACTTCACAACAGAATGGATCAAATCTGTGACGTCCACTTCT 340
             ::::::;:::;:: ||| |:||| |||||||
Db   386 ThrAlaMetLeuAlaAsnLeuPheAlaValProTrpValThrProValThrProLeu 405

Qy   339 CTTTTTGCGGGAAAGTAT------------TATTTTCTCAATGATGGCTGCCC 298
             ::: |||||| :::|||||||| ||||:::
Db   406 IleLeuAlaGlyMetIleLeuGlyProPheLeuGluTrpVal--- 424

Qy   297 TTCAGACAGTGGTTTCTT---------------------------------- 280
             |||||||||
Db   425 ------TrpTyrLeuThrAspArgAlaLeuAlaLeuAlaLeuPheTyrLeuLeuAsnAla 441

Qy   279 ------CAGCTCGTGTGACAAATTAGGAGATGGAATTCTG---ATGTCGTACCTAAGGCTT 230
             ||| :::||| ::: |||
Db   442 LeuProGlnGlyTrpValAsnIleAspAsnArgTrpGlnTrpLeuThrLeuLeuProTrp 461

Qy   229 GTCCATCTTTGTTCGTTGGAGGTTGACACCATTTCTTCTTGGGACGTCCAGATTCCAGCAGCC 170
             ::: |||||| :::|||||
Db   462 LeuThrLeuIleAlaTrpArgLeuAsn-----ValTrpArgThrTrp---ProAlaVal 478

Qy   169 ACCCTCCTCGGGGTGGTGGTGAATGTCTTCAGGCATGGGCAATATGACACAGCAAGTGTT 110
             ::: ||||||| ::||||| ||
Db   479 CysLeuCysGlyLeuLeuMetSerTrpProLeuTrpProArgProIleAsnAlaSerGly 498

Qy   109 TTCCAAATCGGGACCACATCCCCTCAGCTCTGCTGCTGCTAGCGTTCTTCTAAGGAGTTGGC 50
             :::|| |::::
Db   499 TrpGlnVal------HisMetLeuAspVal------GlyGlnGly 509

Qy   49 TCAGCAGCCAGATTGTCTGCAGATAGG------------ACCGGAACAGGCTGG 5
             |:|| ||| ||| || |||||
Db   510 LeuAlaIleAlaIleValArgGlyAspLysValIleLeuTyrAspThrGlyArgAlaTrp 529

Qy   4 CCA 2
           |||
Db   530 Pro 530

RESULT 8

Q8TC33
ID   Q8TC33   PRELIMINARY;   PRT;   411 AA.
AC   Q8TC33;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Similar to hypothetical protein FLJ11220.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis.;
RA   Strausberg R.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC026155; AAH26155.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE 411 AA; 45706 MW; 402A9998904FC210 CRC64;

Alignment Scores:
Pred. No.:   0.428   Length: 411
Score: 88.50   Matches: 39
Percent Similarity: 44.62%   Conservative: 19
Best Local Similarity: 30.00%   Mismatches: 50
Query Match: 11.43%   Indels: 22
DB: 4   Gaps: 7

US-09-518-842-1 (1-420) x Q8TC33 (1-411)

Qy   9 CCTGTTCCGGTCCTATCTGCCAGCAATCTGCTGCTGCTGAGCCAACTCCTTAGAGAAAG 68
           ||||||: ||| ::: :::
Db   99 ProProProSerLeuAlaProAlaGlyGlyProAlaValAlaAlaProLeuPro------Ala 116

Qy   69 CCTAGCAGCAGAGCTGAGGGGATGTGGTCCCGATTGGAAAAACACTTGCTGTCATATTG 128
           ||||||::: |||| ::|||||||
Db   117 ProSerThrSerAlaLeuPheThrPheSerProLeu-----ThrValSerAlaAlaGly 134

Qy   129 CCC----------CATGCTGAGAAGACATTCACCAACCCCAGGAGGTGGCTGCT 176
           ||| |||
Db   135 ProLysHisGlyHisGluArgHisHisHisHisHisHisArg-Gly------ 151

Qy   177 GGAATCTGGACGTCCCAAAGAAATGGTGTCAACCTCCAACACAAAGATGACAAGCCTT 236
           |:: ||||
Db   152 -ProAspGlyAspProSerCysGlyThrAspLeuLysHisLysAspLysGlnGluAs 171

Qy   237 AGGTACGACATCAGAATTCATTCCTAATTG--TCACCAGAGCTGAAGAAACCACTGTC 293
           ::: ||| ::: |||
Db   171 nGlyGluArgThrGlyValProLeuIleLysAlaProLysAlaProGluThrProAspGl 191

Qy   294 TGAAGGGCAGCCATCA------------TTGAAGAAAATAATACTTTCCCGCAA 335
           ||| :::|||||||
Db   191 uAsnGlyLysThrGlnArgAlaAspAspPheValLeuLysLysIle------LysLy 208

Qy   336 AAAGAGAAGTGGACGTCACAGATTGAT 363
           ||||::: :::|||||
Db   208 sLysLysLysLysHisArgGluAsp 217

RESULT 9

```
Q9NUP6
ID   Q9NUP6              PRELIMINARY;      PRT;    754 AA.
AC   Q9NUP6;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein FLJ11220.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA   Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo Y., Sugawara M.,
RA   Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA   Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA   Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA   Ninomiya K., Iwayanagi T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK002082; BAA92075.1; -.
CC   Hypothetical protein.
SQ   SEQUENCE   754 AA;   85004 MW;   596632 3DAE4F5B67 CRC64;

Alignment Scores:
Pred. No.:                        0.494
Score:                           88.50
Percent Similarity:              44.62%          Length:            754
Best Local Similarity:           30.00%          Matches:           39
Query Match:                     11.43%          Conservative:      19
DB:                                              Mismatches:        50
                                                 Indels:            22
                                                 Gaps:              7

S-09_518-842-1 (1-420) x Q9NUP6 (1-754)

Qy    9 CCTGTTCCGGTCCATCTGCCAGAATCTGGCTGCTGCTGAGCCAACTCCTTAGAGAAAG    68
        ||      ||      :: :             |||  ||
Db   99 ProProProSerLeuAlaProLeuAlaGlyProAlaValAlaAlaAlaProLeuPro------Ala   116

Qy   69 CCTAGCAGCAGCAGAGCTGAGGGGATGGTGGTCCCCGATTTGGAAAACACTTGCTGTCATATTG   128
        ||||||  ||      ::|||||   |||       |||  ||
Db  117 ProSerThrSerAlaLeuPhePheSerProLeu------ThrValSerAlaAlaGly   134

Qy  129 CCC------------CATGCTGAGAAGACATTCACCACCACCCCAGGAGGTGGCTGCT   176
        |||                   ||| |||
Db  135 ProLysHisLysGlyHisLysGluArgHisHisLysHisHisHisHisArg-Gly--------   151

Qy  177 GGAATCTGGACGTCCCAAAGAAAATGTGTCAACCTCCAACAACAAAGATGGACAAGCCTT   236
        |||||                     ::  |||||  ||
Db  152 -ProAspGlyAspProSerSerCysGlyThrAspLeuLysHisLysAspLysGlnGluAs   171

Qy  237 AGGTACGACACATCAGAATTCATTCCTAATTTG---TCACCAGAGCTGAGAAAAACCACTGTC   293
        |||            :: :||||               ||| |||
Db  171 nGlyGluArgThrGlyGlyValProLeuIleLysAlaProLeuLysArgGluThrProAspGl   191

Qy  294 TGAAGGGCAGCCATCA---------------------TTGAAGAAAATAAATACTTTCCCGCCA   335
        ||                                   |||||||  :||
Db  191 uAsnGlyLysThrGlnArgAlaAspAspPheValLeuLysLysIle--------LysLy   208

Qy  336 AAAGAGAAGTGGACGTCACAGATTTGAT   363
        :|||||     ::|||||  ||
Db  208 sLysLysLysLysHisArgGluAsp   217

RESULT 10
Q9RLW4
ID   Q9RLW4              PRELIMINARY;      PRT;    1218 AA.
AC   Q9RLW4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Virulence protein S.
GN   KVGS.
OS   Klebsiella pneumoniae.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Klebsiella.
OX   NCBI_TaxID=573;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Yang S.L., Chang H.Y., Peng H.L.;
RT   "Identification and characterization of KvgASQR, a two component
RT   signal transduction system, in Klebsiella pneumoniae.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE
CC       KINASES.
DR   EMBL; AJ250891; CAB61240.1; -.
DR   HSSP; P06143; 1AB6.
DR   InterPro; IPR003594; ATPbind_ATPase.
DR   InterPro; IPR004358; Bact_sens_pr_C.
DR   InterPro; IPR003661; His_kinA.
DR   InterPro; IPR005467; His_kinase.
DR   InterPro; IPR000014; PAS_domain.
DR   InterPro; IPR001789; Response_reg.
DR   InterPro; IPR001311; SBP/glu_receptor.
DR   InterPro; IPR001638; SBP_bac_3.
DR   Pfam; PF02518; HATPase_c; 1.
DR   Pfam; PF00512; HisKA; 1.
DR   Pfam; PF00072; response_reg; 1.
DR   Pfam; PF00497; SBP_bac_3; 1.
DR   PRINTS; PR00344; BCTRLSENSOR.
DR   ProDom; PD000039; Response_reg; 1.
DR   SMART; SM00387; HATPase_c; 1.
DR   SMART; SM00388; HisKA; 1.
DR   SMART; SM00091; PAS; 1.
DR   SMART; SM00062; PBPb; 2.
DR   SMART; SM00448; REC; 1.
DR   PROSITE; PS50109; HIS_KIN; 1.
DR   PROSITE; PS50110; RESPONSE_REGULATORY; 1.
KW   Kinase; Phosphorylation; Sensory transduction; Transferase.
SQ   SEQUENCE   1218 AA;   136317 MW;   24F5889F2B9F02C0 CRC64;

Alignment Scores:
Pred. No.:                        0.721
Score:                           87.50
Percent Similarity:              44.95%          Length:            1218
Best Local Similarity:           27.52%          Matches:           30
Query Match:                     11.30%          Conservative:      19
DB:                                              Mismatches:        47
                                                 Indels:            13
                                                 Gaps:              3

US-09-518-842-1 (1-420) x Q9RLW4 (1-1218)

Qy    7 AGCCTGTC-------CGTTCCTATCTGCCAGCAACTCGGCTGCTGCTGAGCCAACTCCTT   60
        ||||||||           ||||                        |||
Db  620 SerLeuPheAspArgArgSerProAlaValHisIlePheProValIleGlnGlnGluMet   639

Qy   61 AGAGAAGCCTAGCAGCAGAGCTG---------------------AGGGGATGT   93
        ::::::|||   |||||                             |||
Db  640 GlnGlnGlyLeuAlaProGluThrValAlaSerHisGlnPheThrLeuAsnAsnGlyAla   659

Qy   94 GGTCCCGATTTGGAAAACACTTGCTGTCATATTGCCCCATGCCTGAGAAGACATTCACC   153
        ||        :::::  ||||         ||||||||:::::|||
Db  660 GluAspArgThrIleLeuHisHisTrpMetThrLeuCysProLeuProLeuProProPro   679

Qy  154 ACCACCCCAGGAGGGTGG-----------CTGCTGGAATCTGGACGTCCCAAAGAAAATGGTGTCA   207
        |||||||  ||||||            |||        ::  |||
Db  680 ValLeuIleCysGlyTrpGlnAspIleThrGluSerArgLysLeuMetGluAlaLeuGln   699

Qy  208 ACCTCCAACACAAAGATGGACAAGCCTTAGTACGACCATCAGAATTCATTCCTAATTTG   267
        ::  |||                 ::::|||          |||  ||:::
Db  700 ValGluLysAspLysAlaIleGluAlaSerArgAlaLysSerArgPheLeuAlaArgMet   719

Qy  268 TCACCAGAGCTGAGAAAAACCACTGTCT   294
        |||    |||  :::||| |||:::|||
Db  720 SerHisGluIleArgThrProValSer   728
```

RESULT 11
Q8CIS1

```
ID   Q8CIS1        PRELIMINARY;   PRT;   685 AA.
AC   Q8CIS1;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Winged-helix repressor FOXP4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6;
RA   Morrisey E.;
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY135029; AAN08624.1; -.
SQ   SEQUENCE 685 AA; 74031 MW; E9A241A675D2B648 CRC64;
```

Alignment Scores:

```
Pred. No.:              1.22
Score:                 85.00
Percent Similarity:    39.58%
Best Local Similarity: 27.08%
Query Match:           10.98%
DB:                    11
```

US-09-518-842-1 (1-420) x Q8CIS1 (1-685)

```
Qy   46 CTGAGCCAACTCCTTAGAGAAAGCCTAGCAGCA------GAGCTGAGGGGATGTGTCCC  99
            ||| ||||:::||| |||                 :::   ||||||
Db  200 LeuLeuAsnLeuGlnArgGlnGlyLeuValSerLeuGlnProSerGlnAlaSerGlyPro 219

Qy  100 CGATTTGGAAACACTTGCTGTCATATTGCCCATGCCTGAGAAGACATTCACCACCACC  159
Db  220 LeuGlnAlaLeuProGlnAlaAlaLeuAlaValCysPro----------ThrAspLeu  233

Qy  160 CCAGGAGGTGGCTGCTGGAATCTGGACGTCCCAAAGAAATGTGTCAACCTCCAACAAC  219
            ||| ||||| |||        :||   ||
Db  234 ProGlnLeuTrpLysGlyGlu---GlyAlaProGlyGlnProAlaGluAspSerGlyArg 252

Qy  220 AAAGATGGACAAGCCTAGGTACGACA-----------TCAGAATTCATT-----CCT  261
            ::::|||||| :::|||                    :::   |||        |||
Db  253 GlnGluGlyLeuAspLeuAlaSerThrAlaValThrAlaThrSerPheAlaSerProPro 272

Qy  262 AATTTGTCACCAGAGCTGAAG--AAACCACTGTCTGAAGGGCAGCCATCATTGAAGAAA 318
            |||||||
Db  273 LysValSerProProLeuSerHisHisProLeuProAsnGlyGlnProThr--------- 289

Qy  319 ATAATACTTTCCCGCAAAAAGAGAAGTGGACGTCACAGATTTGATCCATTC-------- 369
            ::::||||:::  ::::::            |||    ||
Db  290 ---ValLeuThrSerArgArgAspSerSerHisGluGluThrProSerHisPro      308

Qy  370 ----------------TGT----------TGTGAAGTAATTTGTGACGAT---GGA    396
                            |||          ||||||:::  |||:::||
Db  309 LeuTyrGlyHisGlyGluGluCysTrpProGlyLysTrpProGlyCysGluGlyLeuGly  328

Qy  397 ACTTCAGTTAAA  408
            :::|||
Db  329 GlnPheIleLys 332
```

RESULT 12
Q8CID1

```
ID   Q8CID1        PRELIMINARY;   PRT;   612 AA.
AC   Q8CID1;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Similar to RIKEN cDNA 4921517J23 gene.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N; TISSUE=Kidney;
RA   Strausberg R.;
RL   Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC031386; AAH31386.1; -.
SQ   SEQUENCE 612 AA; 66640 MW; 6F04A05490F64947 CRC64;
```

Alignment Scores:

```
Pred. No.:              1.54
Score:                 84.00
Percent Similarity:    41.67%
Best Local Similarity: 25.00%
Query Match:           10.85%
DB:                    11
```

US-09-518-842-1 (1-420) x Q8CID1 (1-612)

```
Qy   15 CCGGTCCTATCTGCCAGCAATCTGGCTGTCTGCTGAG--------------------CCA  53
            ||| :::                  ||:::                             ||
Db  110 ProGlnValProAlaTyrAsnLeuHisSerProGlnIleSerArgGlyAspValSerPro 129

Qy   54 ACTCCTTAGAGAAAGCCTAGCAGCAGCAGGGGGATGTGAGGGGATGGTGTCCCGATTTGGAAACA 113
            ||||||            :::|||||:::|||
Db  130 ThrProIleTyrLeuArgArgAlaArgAlaGlnGly----------IleValLysGlu  145

Qy  114 CTTGCTGTCATATTGCCCAT-------------GCCTGAGAAGACATTCACCACCAC  158
            ||||||                          :::   |||||
Db  146 IleProLeuTyrLeuProHisSerProMetLeuGluSerThrGluAspCys-LeuValGl 165

Qy  159 CCCAGGAGGTGGCTGCTGGAATCTGGACGTCCCAAAGAAATGTGTCAACCTCCAACAA 218
            ||||||               :::|||                    :::|||
Db  165 uProGlyGluSerLeuArgSer----------ProGluGlnIleSerSerGluGlyGlyCy 183

Qy  219 CAAAGATGGACAAGCCTAGGTACGACA------------------------------TC 248
            :::|||||              :::|||
Db  183 sGlnGluAlaArgAlaLeuGlyAsnThrArgSerIleGlnHisProLeuLeuGlyLysAs 203

Qy  249 AGAATTCATTCCTAATTTGTCA---------------------------------- 270
            |||||||               :::
Db  203 pGlnThrTyrLeuProAsnGlnAsnLeuAlaLeuAlaLeuGlyAlaAlaGlyAsnProGlyAspProAr 223

Qy  271 ------------------------------CCAGAGCTGAAGAAACCACTGTCTGAAGG 299
            :::                          ||||||:::                :::---GluAr
Db  223 gGluGluSerArgMetGlyGlnProGlyGlnProGlyGlyProGlyLeuSerLys---- 240

Qy  300 GCAGCCATCATTGAAGAAAAATAATACTTTCCCGCAAAAAGAAGT  345
            |||                  ::::::|||
Db  240 ---gLysLeuGlyLeuLysLysLysGluLeuValLeuThrGluGlnLeuGlnLysAsn  255
```

RESULT 13
Q9HZB9

```
ID   Q9HZB9        PRELIMINARY;   PRT;   321 AA.
AC   Q9HZB9;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   Hypothetical protein PA3089.
GN   PA3089.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC   Pseudomonadaceae; Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
```

```
Db          |||      |||      ||||||
    319 ---SerProSerTyrGluProHisLeuPheGlyArgGluPro----------Gln 332
Qy  305 GGCTGCCCTTCAGACAGTGGTTC---   TTCAGCTCTTGGTGACAAATTAGGAATGAAT 252
           ||||||  |||  |||                :::  ||||||||||||||
Db  333 PheCysPro---AspGlnSerPheAlaProLeuLeuSerAspProArgGlnSerGluAsn 351
Qy  251 TCTGATGTCGTACCTAAGGCTTGTCCCATCTTGTTGTGTTGGAGGTTGACACC----- 201
           ||||:::       ||||:::       :::||||||
Db  352 IleAlaValProProGlnThrAlaProSerValGlnGlnIleAspThrHisLeuGln 371
Qy  200 ------------------------ATTTCTTTGGGACGTCCAGATTCCAGCAGC 171
                                 ::|||      ||||||||||||||
Db  372 AsnPheSerLeuMetProProAspThrCysGluAlaLeuAlaArgProAspAlaSerSer 391
Qy  170 CACCCTCCTGGGGTGGTGGTGAATGTCTTCTCAGCATGGGCAATATGACAGCAAGTGT 111
           |||
Db  392 ThrProLeuSerThrProLeuPro------------------------ 399
Qy  110 TTTCCAAATCGGGAGCCACCACCATCCCCTCAGCTCT 78
           ||||||||||  ::||||||||:||:||
Db  400 PheProAsnLeuGlyGlyAsnProMetSerThr 410
```

RESULT 15
O64293

```
ID   O64293;
AC   O64293;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Orf1626 gp.
GN   ORf1626.
OS   Streptococcus thermophilus bacteriophage Sfi19.
OC   Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX   NCBI_TaxID=72638;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98160788; PubMed=9499809;
RA   Desiere F., Lucchini S., Brussow H.;
RT   "Evolution of Streptococcus thermophilus bacteriophage genomes by
RT   modular exchanges followed by point mutations and small deletions and
RT   insertions.";
RL   Virology 241:345-356(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Lucchini S., Desiere F., Brussow H.;
RL   Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF115102; AAC39295.1; -.
DR   InterPro; IPR000189; SLT domain.
DR   InterPro; IPR001680; WD40.
DR   Pfam; PF01464; SLT; 1.
DR   PROSITE; PS00678; WD_REPEATS 1; 2.
SQ   SEQUENCE  1626 AA;  178114 MW;  6067EBA5A0079235 CRC64;
```

Alignment Scores:

```
Pred. No.:                      2.89
Score:                          82.50
Percent Similarity:             36.25%
Best Local Similarity:          28.75%
Query Match:                    10.66%
DB:                             9
```

US-09-518-842-1 (1-420) x O64293 (1-1626)

```
Qy    9 CCTGTTCCG--------------------------GTCCTATCTGCC 29
           ||||||||                          |||:::|||
Db 1074 ProValProGlyArgAsnThrTyrAlaValLeuProAlaGluValLeuAsnAla 1093
Qy   30 AGCAATCTGGCTGCTGCAGCGCAACTCCTTAGAGAAAGCCTAGCAGCAGAGCTGAGGGG 89
           |||  |||||               :::::::
Db 1094 SerGluTrpAlaAlaLeuSerGlyAla-------LysProPheAlaLysGlyThrGly 1110
Qy   90 ATGTGGTCCCCGATTCGGAAAAACACTTGCTGTCGTCATATTGCCCCATGGGCTGAAGACATT 149
```

```
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.,
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
DR   EMBL; AE004733; AAG06477.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE  321 AA;  36640 MW;  3E85FDD3DBF268F CRC64;
```

Alignment Scores:

```
Pred. No.:                      1.72
Score:                          83.00
Percent Similarity:             61.54%
Best Local Similarity:          43.59%
Query Match:                    10.72%
DB:                             16
```

US-09-518-842-1 (1-420) x Q9HZB9 (1-321)

```
Qy   94 GGTCCCGATTTGGAAAACACTGCTGTCAT---------------ATTGCCCCATG 135
           |||||||||::  ||||||::|||  |||       ||  :||||||||
Db  234 GlyProGlnThrGlyGlnHisTrpLeu-HisLeuSerArgGluHisTrpLeuAlaProAl 253

Qy  136 CCTGAGAAGACATTCACCA------CCACCCCAGGAGGTGGCTGCTGGAATCTG 184
           ::|||||||  |||          ||||||||||||||:::
Db  253 aMetArgArgGlnAspProLeuAspProProArgAspGlyAlaCysTrpAsnIle 271
```

RESULT 14
Q9BE45

```
ID   Q9BE45
AC   Q9BE45;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   MAIL.
GN   MAIL.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Holstein; TISSUE=Blood;
RA   Yamaji D., Kitamura H., Morimatsu M., Shiina T., Kanehira K.,
RA   Fujikura D., Saito M.;
RT   "Bos taurus mRNA for MAIL, complete cds.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB058410; BAB39767.1; -.
     HSSP; P25963; INFI.
DR   InterPro; IPR002110; ANK.
DR   Pfam; PF00023; ank; 6.
DR   PRINTS; PR01415; ANKYRIN.
DR   SMART; SM00248; ANK; 6.
DR   PROSITE; PS50088; ANK_REPEAT; 3.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
KW   ANK repeat; Repeat.
SQ   SEQUENCE  718 AA;  78247 MW;  CB8DBF6A994A9E00 CRC64;
```

Alignment Scores:

```
Pred. No.:                      2.38
Score:                          82.50
Percent Similarity:             39.69%
Best Local Similarity:          30.53%
Query Match:                    11.07%
DB:                             6
```

US-09-518-842-1 (1-420) x Q9BE45 (1-718)

```
Qy  419 TATGTACATAAT-----TTAACTGAAGTTCATCGTCACAAATTACTTCAACAGAAT 366
           || |||        |||::  |||||||::
Db  302 TyrThrHisAsnProAlaLeuGluTyrSerProTyrSerArg---ThrSerGln----- 318

Qy  365 GGATCAAATCTGTGACGTCCACTTCTCTTTTTGCGGAAAGTATATTTTTCTCAATGAT 306
```

Db

        ||||| ||||| |||
1111 PheTrpSerLysIleTrpAsnThrThr------------------- 1120

y   150 CACCACCACCCCAGGAGGTGGCTGCTGAATCTGGACGTCCCAAAGAAATGGTGTCAAC 209
                        ||||| :::|||||||||:::
b   1121 --------------AsnValAlaGlySerValTrpAsnGlyLeuLysAsnGlyValAsp 1135

earch completed: October 9, 2003, 12:18:38
ob time : 84 secs

```
            Copyright (c) 1993 - 2003 Compugen Ltd.
                    GenCore version 5.1.6
```

DM nucleic - protein search, using frame_plus_n2p model

Run on: October 9, 2003, 12:13:06 ; Search time 21 Seconds
                                      (without alignments)
                           1692.434 Million cell updates/sec

Title: US-09-518-842-1
Perfect score: 774
Sequence: 1 ATGGCCAGCCTGTTCCGGTC.........CAGTTAAATTATGTACATAG 420

Scoring table: BLOSUM62
        Xgapop 10.0 , Xgapext 0.5
        Ygapop 10.0 , Ygapext 0.5
        Fgapop   6.0 , Fgapext 7.0
        Delop   6.0 , Delext   7.0

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 657434

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
MODELframe+_n2p.model -DEV=xlp
Q=/cgn2_1/USPTO_spool/US09518842/runat_09102003_111038_13105/app_query.fasta_1.583
DB=Issued_Patents_AA -QFMT=fastar -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0
LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
USER=US09518842 @CGN_1_1_38 @runat_09102003_111038_13105 -NCPU=6 -ICPU=3
NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
DEV_TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_AA.*
 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep;*
 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep;*
 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep;*
 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep;*
 5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep;*
 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep;*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Match	Query Match Length	DB	ID	Description
1	739	95.5	139	3	US-08-950-720A-15	Sequence 15, Appl
2	739	95.5	139	3	US-08-991-890-2	Sequence 2, Appli
3	733	94.7	139	4	US-09-599-564A-2	Sequence 2, Appli
4	733	94.7	174	3	US-09-174-465D-2	Sequence 2, Appli
5	620	80.1	124	3	US-08-991-890-5	Sequence 5, Appli
6	619	80.0	159	3	US-08-991-890-4	Sequence 4, Appli
7	255	32.9	51	3	US-09-174-465D-10	Sequence 10, Appl
8	255	32.9	51	4	US-09-599-564A-10	Sequence 10, Appl
9	228	29.5	41	3	US-09-174-465D-8	Sequence 8, Appli
10	228	29.5	41	4	US-09-599-564A-8	Sequence 8, Appli
11	228	29.5	41	4	US-09-201-227A-27	Sequence 27, Appl
12	171	22.1	30	3	US-09-174-465D-12	Sequence 12, Appl
13	171	22.1	30	4	US-09-599-564A-12	Sequence 12, Appl
14	147	19.0	25	4	US-09-201-227A-28	Sequence 28, Appl
15	123.5	16.0	185	3	US-08-950-720A-12	Sequence 12, Appl
16	119.5	15.4	185	3	US-08-950-720A-13	Sequence 13, Appl
17	113.5	14.7	150	5	US-08-443-568B-12	Sequence 12, Appl
18	113.5	14.7	150	1	PCT-US94-06997-12	Patent No. 5464756
19	113.5	14.7	162	6	5464756-18	Patent No. 5464756
20	101.5	13.1	164	6	5464756-20	Sequence 6, Appli
21	85	11.0	17	3	US-09-174-465D-6	Sequence 6, Appli
22	85	11.0	17	4	US-09-599-564A-6	Sequence 28877, A
23	83	10.7	324	6	US-09-252-991A-28877	Patent No. 5212074
24	80	10.3	322	6	5212074-1	Sequence 3, Appli
25	75	10.1	994	4	US-08-699-103B-3	Sequence 3, Appli
26	75	10.1	994	4	US-09-229-059-3	Sequence 3, Appli
27	75	10.1	994	4	US-09-628-133-3	Sequence 15, Appl
28	75	10.1	995	4	US-08-747-562-15	Sequence 2, Appli
29	75	10.1	1711	5	US-08-342-930-2	Sequence 24085, A
30	74	9.6	206	4	US-09-252-991A-24085	Sequence 1, Appli
31	73	9.4	458	4	US-08-618-485B-1	Patent No. 5177002
32	73	9.4	458	6	5177002-1	Sequence 193, App
33	71.5	9.2	778	4	US-09-556-877-193	Sequence 193, App
34	71.5	9.2	778	4	US-09-620-412C-193	Sequence 178, App
35	71.5	9.2	778	4	US-09-598-419-193	Sequence 178, App
36	71.5	9.2	1530	4	US-09-556-877-178	Sequence 178, App
37	71.5	9.2	1530	4	US-09-620-412C-178	Sequence 178, App
38	71.5	9.2	1530	4	US-09-598-419-178	Sequence 3, Appli
39	70	9.0	344	1	US-08-843-993-3	Sequence 3, Appli
40	70	9.0	344	3	US-09-059-520A-3	Sequence 3, Appli
41	70	9.0	344	3	US-09-334-275-3	Sequence 33133, A
42	69.5	9.0	754	4	US-09-252-991A-33133	Sequence 1, Appli
43	69	8.9	239	5	PCT-US93-01652-1	Patent No. 5177002
44	69	8.9	458	6	5177002-2	Sequence 5, Appli
45	69	8.9	474	1	US-08-222-619-5	

ALIGNMENTS

```
RESULT 1
US-08-950-720A-15
; Sequence 15, Application US/08950720A
; Patent No. 6046028
; GENERAL INFORMATION:
;    APPLICANT: Conklin, Darrell C.
;    APPLICANT: Lofton-Day, Catherine E.
;    APPLICANT: Lok, Si
;    APPLICANT: Jaspers, Stephen R.
;    TITLE OF INVENTION: INSULIN HOMOLOG
;    NUMBER OF SEQUENCES: 17
;    CORRESPONDENCE ADDRESS:
;       ADDRESSEE: ZymoGenetics, Inc.
;       STREET: 1201 Eastlake Avenue East
;       CITY: Seattle
;       STATE: WA
;       COUNTRY: USA
;       ZIP: 98102
;    COMPUTER READABLE FORM:
;       MEDIUM TYPE: Diskette
;       COMPUTER: IBM Compatible
;       OPERATING SYSTEM: DOS
;       SOFTWARE: FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/08/950,720A
;       FILING DATE:
;       CLASSIFICATION: 435
;    PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;       NAME: Sawislak, Deborah A
;       REGISTRATION NUMBER: 37,438
;       REFERENCE/DOCKET NUMBER: 96-09
;    TELECOMMUNICATION INFORMATION:
;       TELEPHONE: 206-442-6672
```

;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/991,890
;   FILING DATE:
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 60/033,003
;   FILING DATE: December 16, 1996
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Sawislak, Deborah A
;   REGISTRATION NUMBER: 37,438
;   REFERENCE/DOCKET NUMBER: 96-41
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 206-442-6672
;   TELEFAX: 206-442-6678
;   TELEX:
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 139 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   FRAGMENT TYPE: internal
; US-08-991-890-2

Alignment Scores:
Pred. No.:
Score: 4.05e-84   Length: 139
Percent Similarity: 739.00   Matches: 139
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 100.00%   Mismatches: 0
DB: 95.48%   Indels: 0
                                    Gaps: 0

US-09-518-842-1 (1-420) x US-08-991-890-2 (1-139)

Qy   1 ATGGCCAGCCTGTTCCGGTCCTATCTGCCAGAATCTGGCTGCTGCTGAGCCAACTCCTT   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MetAlaSerLeuPheArgSerTyrLeuProAlaIleTrpLeuLeuLeuSerGlnLeuLeu   20

Qy   61 AGAGAAAGCCTAGCAGCAGAGCTGAGGGGATGGTGCCCGATTGGAAAAACACTTGCTG   120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21 ArgGluSerLeuAlaAlaGluLeuArgGlyCysGlyProArgPheGlyLysHisLeuLeu   40

Qy   121 TCATATTGCCCATGCCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAA   180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41 SerTyrCysProMetProGluLysThrPheThrThrThrProGlyGlyTrpLeuLeuGlu   60

Qy   181 TCTGGACGTCCCAAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGT   240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 SerGlyArgProLysGluMetValSerThrSerAsnAsnLysAspGlyGlnAlaLeuGly   80

Qy   241 ACGACATCAGAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGG   300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81 ThrThrSerGluPheIleProAsnLeuSerProGluLeuLysLysProLeuSerGluGly   100

Qy   301 CAGCCATCATTGAAGAAATAATACTTTCCCGCAAAAGAAGAAGTGACGTCACAGATTT   360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   101 GlnProSerLeuLysIleIleLeuSerArgLysLysLysArgSerGlyArgHisArgPhe   120

Qy   361 GATCCATTCTGTTGTGAAGTAATTTGTGACGATGGAACTTCAGTAAATTATGTACA   417
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AspProPheCysCysGluValIleCysAspAspGlyThrSerValLysLeuCysThr   139

RESULT 3
US-09-599-564A-2
; Sequence 2, Application US/09599564A
; Patent No. 6362318
; GENERAL INFORMATION:
; APPLICANT: KOMAN, Ahment
; APPLICANT: CHASSIN, Dorine
; APPLICANT: BELLET, Dominique

Left column

```
, TITLE OF INVENTION:    NEW PROTEIN CALLED EPIL/PLACENTIN, PROCESS FOR THE
,, TITLE OF INVENTION:   PREPARATION OF THIS PROTEIN AND PHARMACEUTICAL
,  TITLE OF INVENTION:   COMPOSITION CONTAINING SUCH, DNA CODING FOR SAID
   TITLE OF INVENTION:   PROTEIN
   FILE REFERENCE:  017753-127
   CURRENT APPLICATION NUMBER: US/09/599,564A
   CURRENT FILING DATE: 2000-06-23
   PRIOR APPLICATION NUMBER: 09/174,465
   PRIOR FILING DATE: 1998-10-19
   PRIOR APPLICATION NUMBER: US 08/482,842
   PRIOR FILING DATE: 1995-06-07
   NUMBER OF SEQ ID NOS: 16
   SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 2
   LENGTH: 139
   TYPE: PRT
   ORGANISM: Unknown
   FEATURE:
   OTHER INFORMATION: Description of Unknown Organism:EPIL - Early
   OTHER INFORMATION: Placenta Insulin-Like Peptide
US-09-599-564A-2

Alignment Scores:
Pred. No.:
Score:                  2.29e-83
                        733.00
Percent Similarity:     99.28%          Length:          139
Best Local Similarity:  99.28%          Matches:         138
Query Match:            94.70%          Conservative:    0
DB:                                     Mismatches:      1
                                        Indels:          1
                                        Gaps:            0

US-09-518-842-1 (1-420) x US-09-599-564A-2 (1-139)

Qy   1 ATGGCCAGCCTGTTCCGGTCCTATCTGCCAGCAATCTGGTGCTGCTGAGCCAACTCCTT 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MetAlaSerLeuPheArgSerTyrLeuProAlaIleTrpLeuLeuLeuSerGlnLeuLeu 20

Qy  61 AGAGAAAGCCTAGCAGCAGAGCTGAGGGGATGTGGTCCCGATTTGGAAAACACTTGCTG 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21 ArgGluSerLeuAlaAlaGluLeuArgGlyCysGlyProArgPheGlyLysHisLeuLeu 40

Qy 121 TCATATTGCCCCATGCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAA 180
        ||||||||||||||||||||||||||||||||||||||||||||||||
Db  41 SerTyrCysProMetProGluLysThrPheThrThrThrProGlyGlyTrpLeuLeuGlu 60

Qy 181 TCTGGACGTCCCAAAGAAATGGTGTCAACCTCCAACAACAAGATGACAAGCCTTAGGT 240
        ||||||||||||||||||||   |||||||||||||||||||
Db  61 SerGlyArgProLysGluMetValSerThrSerAsnLysAspGlyGlnAlaLeuGly 80

Qy 241 ACGACATCAGAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGG 300
        ||||||||||||||||||||||||||||||||||||||||||||||
Db  81 ThrThrSerGluPheGluProAsnLeuSerProGluLeuLysLysProLeuSerGluGly 100

Qy 301 CAGCCATCATTGAAGAAAATAATACTTTCCCGCAAAAAGAGAAGTGGACGTCACAGATT 360
        |||||||||||||||||||||||||||||||||
Db 101 GlnProSerLeuLysLysIleIleLeuSerArgLysLysArgSerGlyArgHisArgPhe 120

Qy 361 GATCCATTCTGTTGTGAAGTAATTTGTGACGATGAACTTCAGTTAAATTATGTACA 417
        ||||||||||||||||||||
Db 121 AspProPheCysCysGluValIleCysAspAspGlyThrSerValLysLysLeuCysThr 139
```

RESULT 4
S-09-174-465D-2
Sequence 2, Application US/09174465D
Patent No. 6180364
GENERAL INFORMATION:
APPLICANT: KOMAN, Ahment
APPLICANT: CHASSIN, Dorine
APPLICANT: BELLET, Dominique
TITLE OF INVENTION: NEW PROTEIN CALLED EPIL/PLACENTIN, PROCESS FOR THE
TITLE OF INVENTION: PREPARATION OF THIS PROTEIN AND PHARMACEUTICAL
TITLE OF INVENTION: COMPOSITION CONTAINING SUCH, DNA CODING FOR SAID
TITLE OF INVENTION: PROTEIN
FILE REFERENCE: 017753-103

Right column

```
   CURRENT APPLICATION NUMBER: US/09/174,465D
   CURRENT FILING DATE: 1998-10-19
   PRIOR APPLICATION NUMBER: US 08/482,842
   PRIOR FILING DATE: 1995-06-07
   NUMBER OF SEQ ID NOS: 16
   SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 2
   LENGTH: 174
   TYPE: PRT
   ORGANISM: Unknown
   FEATURE:
   OTHER INFORMATION: Description of Unknown Organism:EPIL - Early
   OTHER INFORMATION: Placenta Insulin-Like Peptide
US-09-174-465D-2

Alignment Scores:
Pred. No.:                              Length:          174
Score:                  2.56e-83        Matches:         138
Percent Similarity:     733.00          Conservative:    0
Best Local Similarity:  99.28%          Mismatches:      1
Query Match:            99.28%          Indels:          0
DB:                     94.70%          Gaps:            0

US-09-518-842-1 (1-420) x US-09-174-465D-2 (1-174)

Qy   1 ATGGCCAGCCTGTTCCGGTCCTATCTGCCAGCAATCTGGTGCTGCTGAGCCAACTCCTT 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36 MetAlaSerLeuPheArgSerTyrLeuProAlaIleTrpLeuLeuLeuSerGlnLeuLeu 55

Qy  61 AGAGAAAGCCTAGCAGCAGAGCTGAGGGGATGTGGTCCCGATTTGGAAAACACTTGCTG 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  56 ArgGluSerLeuAlaAlaGluLeuArgGlyCysGlyProArgPheGlyLysHisLeuLeu 75

Qy 121 TCATATTGCCCCATGCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAA 180
        ||||||||||||||||||||||||||||||||||||||||||||||||
Db  76 SerTyrCysProMetProGluLysThrPheThrThrThrProGlyGlyTrpLeuLeuGlu 95

Qy 181 TCTGGACGTCCCAAAGAAATGGTGTCAACCTCCAACAACAAGATGACAAGCCTTAGGT 240
        ||||||||||||||||||||   |||||||||||||||||||
Db  96 SerGlyArgProLysGluMetValSerThrSerAsnLysAspGlyGlnAlaLeuGly 115

Qy 241 ACGACATCAGAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGG 300
        ||||||||||||||||||||||||||||||||||||||||||||||
Db 116 ThrThrSerGluPheGluProAsnLeuSerProGluLeuLysLysProLeuSerGluGly 135

Qy 301 CAGCCATCATTGAAGAAAATAATACTTTCCCGCAAAAAGAGAAGTGGACGTCACAGATT 360
        |||||||||||||||||||||||||||||||||
Db 136 GlnProSerLeuLysLysIleIleLeuSerArgLysLysArgSerGlyArgHisArgPhe 155

Qy 361 GATCCATTCTGTTGTGAAGTAATTTGTGACGATGAACTTCAGTTAAATTATGTACA 417
        ||||||||||||||||||||
Db 156 AspProPheCysCysGluValIleCysAspAspGlyThrSerValLysLysLeuCysThr 174
```

RESULT 5
US-08-991-890-5
; Sequence 5, Application US/08991890
; Patent No. 6114307
GENERAL INFORMATION:
   APPLICANT: Jaspers, Stephen R.
   APPLICANT: Sprugel, Katherine H.
   APPLICANT: Ren, Hong Ping
   APPLICANT: Humes, Jacqueline M.
   APPLICANT: Hofman, Ross C.
   APPLICANT: Conklin, Darrell C.
   TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
   TITLE OF INVENTION: STIMULATING PANCREATIC ISLET CELL REGENERATION
   NUMBER OF SEQUENCES: 7
   CORRESPONDENCE ADDRESS:
   ADDRESSEE: ZymoGenetics, Inc.
   STREET: 1201 Eastlake Avenue East
   CITY: Seattle
   STATE: WA
   COUNTRY: USA

APPLICANT: Conklin, Darrell C.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
TITLE OF INVENTION: STIMULATING PANCREATIC ISLET CELL REGENERATION
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
  ADDRESSEE: ZymoGenetics, Inc.
  STREET: 1201 Eastlake Avenue East
  CITY: Seattle
  STATE: WA
  COUNTRY: USA
  ZIP: 98102
COMPUTER READABLE FORM:
  MEDIUM TYPE: Diskette
  COMPUTER: IBM Compatible
  OPERATING SYSTEM: DOS
  SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/991,890
  FILING DATE:
  CLASSIFICATION:
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: 60/033,003
  FILING DATE: December 16, 1996
ATTORNEY/AGENT INFORMATION:
  NAME: Sawislak, Deborah A
  REGISTRATION NUMBER: 37,438
  REFERENCE/DOCKET NUMBER: 96-41
TELECOMMUNICATION INFORMATION:
  TELEPHONE: 206-442-6672
  TELEFAX: 206-442-6678
  TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
  LENGTH: 159 amino acids
  TYPE: amino acid
  STRANDEDNESS: single
  TOPOLOGY: linear
  MOLECULE TYPE: protein
  FRAGMENT TYPE: internal

US-08-991-890-4

Alignment Scores:
Pred. No.:   4:
Score: 4.86e-69   Length: 159
Percent Similarity: 619.00   Matches: 114
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 100.00%   Mismatches: 0
DB: 79.97%   Indels: 0
                            3   Gaps: 0

US-09-518-842-1 (1-420) x US-08-991-890-4 (1-159)

Qy   76 GCAGAGCTGAGGGGATGGTCCCGATTTGGAAAACACTTGCTGTCATATTGCCCATG 135
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   46 AlaGluLeuArgGlyCysGlyProArgPheGlyLysHisLeuLeuSerTyrCysProMet 65

Qy   136 CCTGAGAAGACATTCACCACCACCCAGGAGGTGGCTGCTGGAATTGGACGTCCCAAA 195
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   66 ProGluLysThrPheThrThrThrProGlyGlyTrpLeuLeuGluSerGlyArgProLys 85

Qy   196 GAAATGGTGTCAACCTCCAACAACAAGATGACAAGCCTTAGGTACGACACATCAGAATTC 255
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   86 GluMetValSerThrSerAsnAsnLysAspGlyGlnAlaLeuGlyThrThrSerGluPhe 105

Qy   256 ATTCCTAATTTGTCACCAGAGTCTGAAGAAACCACTGTCTGAAGGGCAGCCATCATTGAAG 315
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   106 IleProAsnLeuSerProLeuLysProLeuLysLysProLeuSerGluGluGlyGlnProSerLeuLys 125

Qy   316 AAAATAATACTTTCCCGCAAAAAGAAGAAGTGGACGTCACAGATTTGATCCATTCTGTTGT 375
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   126 LysIleIleLeuSerArgLysLysArgGlyArgHisArgPheAspProPheCysCys 145

Qy   376 GAAGTAATTGTGACGATGGAACTTCAGTTAAATTATGTACA 417
         |||||||||||||||||||||||||||||||||||||||||

ZIP: 98102
COMPUTER READABLE FORM:
  MEDIUM TYPE: Diskette
  COMPUTER: IBM Compatible
  OPERATING SYSTEM: DOS
  SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/991,890
  FILING DATE:
  CLASSIFICATION:
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: 60/033,003
  FILING DATE: December 16, 1996
ATTORNEY/AGENT INFORMATION:
  NAME: Sawislak, Deborah A
  REGISTRATION NUMBER: 37,438
  REFERENCE/DOCKET NUMBER: 96-41
TELECOMMUNICATION INFORMATION:
  TELEPHONE: 206-442-6672
  TELEFAX: 206-442-6678
  TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
  LENGTH: 124 amino acids
  TYPE: amino acid
  STRANDEDNESS: single
  TOPOLOGY: linear

US-08-991-890-5

Alignment Scores:
Pred. No.:   5:
Score: 3.23e-69   Length: 124
Percent Similarity: 620.00   Matches: 114
Best Local Similarity: 100.00%   Conservative: 1
Query Match: 99.13%   Mismatches: 0
DB: 80.10%   Indels: 0
                            3   Gaps: 0

US-09-518-842-1 (1-420) x US-08-991-890-5 (1-124)

Qy   73 GCAGCAGAGCTGAGGGGATGGTCCCGATTTGGAAAACACTTGCTGTCATATTGCCCC 132
         ::|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10 SerAlaGluLeuArgGlyCysGlyProArgPheGlyLysHisLeuLeuSerTyrCysPro 29

Qy   133 ATGCCTGAGAAGACATTCACCACCACCCAGGAGGTGGCTGCTGGAATCTGACGTCCC 192
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   30 MetProGluLysThrPheThrThrThrProGlyGlyTrpLeuLeuGluSerGlyArgPro 49

Qy   193 AAAGAAATGGTGTCAACCTCCAACAACAAGATGACAAGCAGCCTTAGGTACGACACATCAGAA 252
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   50 LysGluMetValSerThrSerAsnAsnLysAspGlyGlnAlaLeuGlyThrThrSerGlu 69

Qy   253 TTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCATTG 312
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   70 PheIleProAsnLeuSerProLeuLysProLeuLysLysProLeuSerGluGluGlyGlnProSerLeu 89

Qy   313 AAGAAAATAATACTTTCCCGCAAAAAGAAGAAGTGGACGTCACAGATTTGATCCATTCTGT 372
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   90 LysLysIleIleLeuSerArgLysLysArgGlyArgHisArgPheAspProPheCys 109

Qy   373 TGTGAAGTAATTGTGACGATGGAACTTCAGTTAAATTATGTACA 417
         |||||||||||||||||||||||||||||||||||||
Db   110 CysGluValIleValThrSerValLysLysLeuCysThr 124

RESULT 6
US-08-991-890-4
, Sequence 4, Application US/08991890
, Patent No. 6114307
GENERAL INFORMATION:
  APPLICANT: Jaspers, Stephen R.
  APPLICANT: Sprugel, Katherine H.
  APPLICANT: Ren, Hong Ping
  APPLICANT: Humes, Jacqueline M.
  APPLICANT: Hoffman, Ross C.

Db   146 GluValIleCysAspGlyThrSerValLysLeuCysThr 159

RESULT 7
US-09-174-465D-10
; Sequence 10, Application US/09174465D
Patent No. 6180364
GENERAL INFORMATION:
APPLICANT: KOMAN, Ahment
APPLICANT: CHASSIN, Dorine
APPLICANT: BELLET, Dominique
TITLE OF INVENTION: NEW PROTEIN CALLED EPIL/PLACENTIN, PROCESS FOR THE
TITLE OF INVENTION: PREPARATION OF THIS PROTEIN AND PHARMACEUTICAL
TITLE OF INVENTION: COMPOSITION CONTAINING SUCH, DNA CODING FOR SAID
TITLE OF INVENTION: PROTEIN
FILE REFERENCE: 017753-103
CURRENT APPLICATION NUMBER: US/09/174,465D
CURRENT FILING DATE: 1998-10-19
PRIOR APPLICATION NUMBER: US 08/482,842
PRIOR FILING DATE: 1995-06-07
NUMBER OF SEQ ID NOS: 16
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 10
  LENGTH: 51
  TYPE: PRT
  ORGANISM: Unknown
  FEATURE:
  OTHER INFORMATION: Description of Unknown Organism:EPIL - Early
  OTHER INFORMATION: Placenta Insulin-Like Peptide
US-09-174-465D-10

Alignment Scores:
Pred. No.:   1.27e-23   Length: 51
Score: 255.00   Matches: 51
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 32.95%   Indels: 0
DB: 3   Gaps: 0

US-09-518-842-1 (1-420) x US-09-174-465D-10 (1-51)

Qy   175 CTGGAATCTGACGTCCAAAGAAATGGTGTCAACCTCCAACACAAAGATGGACAAGCC 234
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 LeuGluSerGlyArgProLysGluMetValSerThrSerAsnLysAspGlyGlnAla 20

Qy   235 TTAGGTACGACATCAGAATTCATTCACCAGAGCTGAAGAAACCACTGTCT 294
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21 LeuGlyThrThrSerGluPheIleProAsnLeuSerProGluLeuLysLysProLeuSer 40

Qy   295 GAAGGGCAGCCATCATTGAAGAAATAATACTT 327
          ||||||||||||||||||||||||||||||||
Db   41 GluGlyGlnProSerLeuLysLysIleIleLeu 51

US-09-174-465D-10

RESULT 8
US-09-599-564A-10
Sequence 10, Application US/09599564A
Patent No. 6362318
GENERAL INFORMATION:
APPLICANT: KOMAN, Ahment
APPLICANT: CHASSIN, Dorine
APPLICANT: BELLET, Dominique
TITLE OF INVENTION: NEW PROTEIN CALLED EPIL/PLACENTIN, PROCESS FOR THE
TITLE OF INVENTION: PREPARATION OF THIS PROTEIN AND PHARMACEUTICAL
TITLE OF INVENTION: COMPOSITION CONTAINING SUCH, DNA CODING FOR SAID
TITLE OF INVENTION: PROTEIN
FILE REFERENCE: 017753-127
CURRENT APPLICATION NUMBER: US/09/599,564A
CURRENT FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: 09/174,465
PRIOR FILING DATE: 1998-10-19
PRIOR APPLICATION NUMBER: US 08/482,842
PRIOR FILING DATE: 1995-06-07
NUMBER OF SEQ ID NOS: 16
SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 10
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism:EPIL - Early
; OTHER INFORMATION: Placenta Insulin-Like Peptide
US-09-599-564A-10

Alignment Scores:
Pred. No.:   1.27e-23   Length: 51
Score: 255.00   Matches: 51
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 32.95%   Indels: 0
DB: 4   Gaps: 0

US-09-518-842-1 (1-420) x US-09-599-564A-10 (1-51)

Qy   175 CTGGAATCTGGACGTCCAAAGAAATGGTGTCAACCTCCAACACAAAGATGACAAGCC 234
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 LeuGluSerGlyArgProLysGluMetValSerThrSerAsnLysAspGlyGlnAla 20

Qy   235 TTAGGTACGACATCAGAATTCATTCCTAATTGTCACCAGAGCTGAAGAAACCACTGTCT 294
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21 LeuGlyThrThrSerGluPheIleProAsnLeuSerProGluLeuLysLysProLeuSer 40

Qy   295 GAAGGGCAGCCATCATTGAAGAAATAATACTT 327
          ||||||||||||||||||||||||||||||||
Db   41 GluGlyGlnProSerLeuLysLysIleIleLeu 51

US-09-518-842-1 (1-420) x US-09-599-564A-10 (1-51)

RESULT 9
US-09-174-465D-8
; Sequence 8, Application US/09174465D
Patent No. 6180364
GENERAL INFORMATION:
APPLICANT: KOMAN, Ahment
APPLICANT: CHASSIN, Dorine
APPLICANT: BELLET, Dominique
TITLE OF INVENTION: NEW PROTEIN CALLED EPIL/PLACENTIN, PROCESS FOR THE
TITLE OF INVENTION: PREPARATION OF THIS PROTEIN AND PHARMACEUTICAL
TITLE OF INVENTION: COMPOSITION CONTAINING SUCH, DNA CODING FOR SAID
TITLE OF INVENTION: PROTEIN
FILE REFERENCE: 017753-103
CURRENT APPLICATION NUMBER: US/09/174,465D
CURRENT FILING DATE: 1998-10-19
PRIOR APPLICATION NUMBER: US 08/482,842
PRIOR FILING DATE: 1995-06-07
NUMBER OF SEQ ID NOS: 16
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 8
  LENGTH: 41
  TYPE: PRT
  ORGANISM: Unknown
  FEATURE:
  OTHER INFORMATION: Description of Unknown Organism:EPIL - Early
  OTHER INFORMATION: Placenta Insulin-Like Peptide
US-09-174-465D-8

Alignment Scores:
Pred. No.:   2.77e-20   Length: 41
Score: 228.00   Matches: 41
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 29.46%   Indels: 0
DB: 3   Gaps: 0

US-09-518-842-1 (1-420) x US-09-174-465D-8 (1-41)

Qy   52 CAACTCCTTAGAGAAGCCTAGCAGCAGAGCTGAGGGGATGTGGTCCCGATTTGGAAAA 111
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GlnLeuLeuArgGluSerLeuAlaAlaGluLeuArgGlyCysGlyProArgPheGlyLys 20

```
;   NUMBER OF SEQ ID NOS: 45
;   SOFTWARE: PatentIn version 3.0
;   SEQ ID NO 27
;   LENGTH: 41
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-201-227A-27

Alignment Scores:
Pred. No.:                2.77e-20       Length:          41
Score:                    228.00         Matches:         41
Percent Similarity:       100.00%        Conservative:     0
Best Local Similarity:    100.00%        Mismatches:       0
Query Match:              29.46%         Indels:           0
DB:                       4              Gaps:             0

US-09-518-842-1 (1-420) x US-09-201-227A-27  (1-41)

Qy     52  CAACTCCTTAGAGAAAGCTAGCAGCAGCTGAGCGGATGTGGTCCCGATTTGGAAAA  111
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  GlnLeuLeuArgGluSerLeuAlaAlaGluAlaGluArgGlyCysGlyProArgPheGlyLys  20

Qy    112  CACTTGCTGTCATATTGCCCCATGCCCGAGAAGACATTCACCACCACCCAGGAGGGTGG  171
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21  HisLeuLeuSerTyrCysProMetProGluLysThrPheThrThrThrProGlyGlyTrp  40

Qy    172  CTG  174
           |||
Db     41  Leu  41
```

```
RESULT 12
US-09-174-465D-12
;   Sequence 12, Application US/09174465D
;   Patent No. 6180364
;   GENERAL INFORMATION:
;   APPLICANT: KOMAN, Ahment
;   APPLICANT: CHASSIN, Dorine
;   APPLICANT: BELLET, Dominique
;   TITLE OF INVENTION: NEW PROTEIN CALLED EPIL/PLACENTIN, PROCESS FOR THE
;   TITLE OF INVENTION: PREPARATION OF THIS PROTEIN AND PHARMACEUTICAL
;   TITLE OF INVENTION: COMPOSITION CONTAINING SUCH, DNA CODING FOR SAID
;   TITLE OF INVENTION: PROTEIN
;   FILE REFERENCE: 017753-103
;   CURRENT APPLICATION NUMBER: US/09/174,465D
;   CURRENT FILING DATE: 1998-10-19
;   PRIOR APPLICATION NUMBER: US 08/482,842
;   PRIOR FILING DATE: 1995-06-07
;   NUMBER OF SEQ ID NOS: 16
;   SOFTWARE: PatentIn Ver. 2.0
;   SEQ ID NO 12
;   LENGTH: 30
;   TYPE: PRT
;   ORGANISM: Unknown
;   FEATURE:
;   OTHER INFORMATION: Description of Unknown Organism:EPIL - Early
;   OTHER INFORMATION: Placenta Insulin-Like Peptide
US-09-174-465D-12

Alignment Scores:
Pred. No.:                3.36e-13       Length:          30
Score:                    171.00         Matches:         30
Percent Similarity:       100.00%        Conservative:     0
Best Local Similarity:    100.00%        Mismatches:       0
Query Match:              22.09%         Indels:           0
DB:                       3              Gaps:             0

US-09-518-842-1 (1-420) x US-09-174-465D-12  (1-30)

Qy    328  TCCCGCAAAAAGAAGAGAAGTGGACTCAACAGATTGATCCATTCTTCTTGTTGTGAAGTAAATTGT  387
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  SerArgLysLysArgSerGlyArgHisArgPheAspProPheCysCysGluValIleCys  20

Qy    388  GACGATGGAACTTCAGTTAAATTATGTACA  417
```

(Left column — rotated text continued)

```
Qy    112  CACTTGCTGTCATATTGCCCCATGCCCGAGAAGACATTCACCACCACCCAGGAGGGTGG  171
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21  HisLeuLeuSerTyrCysProMetProGluLysThrPheThrThrThrProGlyGlyTrp  40

Qy    172  CTG  174
           |||
Db     41  Leu  41
```

```
RESULT 10
US-09-599-564A-8
;   Sequence 8, Application US/09599564A
;   Patent No. 6362318
;   GENERAL INFORMATION:
;   APPLICANT: KOMAN, Ahment
;   APPLICANT: CHASSIN, Dorine
;   APPLICANT: BELLET, Dominique
;   TITLE OF INVENTION: NEW PROTEIN CALLED EPIL/PLACENTIN, PROCESS FOR THE
;   TITLE OF INVENTION: PREPARATION OF THIS PROTEIN AND PHARMACEUTICAL
;   TITLE OF INVENTION: COMPOSITION CONTAINING SUCH, DNA CODING FOR SAID
;   TITLE OF INVENTION: PROTEIN
;   FILE REFERENCE: 017753-127
;   CURRENT APPLICATION NUMBER: US/09/599,564A
;   CURRENT FILING DATE: 2000-06-23
;   PRIOR APPLICATION NUMBER: 09/174,465
;   PRIOR FILING DATE: 1998-10-19
;   PRIOR APPLICATION NUMBER: US 08/482,842
;   PRIOR FILING DATE: 1995-06-07
;   NUMBER OF SEQ ID NOS: 16
;   SOFTWARE: PatentIn Ver. 2.0
;   SEQ ID NO 8
;   LENGTH: 41
;   TYPE: PRT
;   ORGANISM: Unknown
;   FEATURE:
;   OTHER INFORMATION: Description of Unknown Organism:EPIL - Early
;   OTHER INFORMATION: Placenta Insulin-Like Peptide
US-09-599-564A-8

Alignment Scores:
Pred. No.:                2.77e-20       Length:          41
Score:                    228.00         Matches:         41
Percent Similarity:       100.00%        Conservative:     0
Best Local Similarity:    100.00%        Mismatches:       0
Query Match:              29.46%         Indels:           0
DB:                       4              Gaps:             0

US-09-518-842-1 (1-420) x US-09-599-564A-8  (1-41)

Qy     52  CAACTCCTTAGAGAAAGCTAGCAGCAGCTGAGCGGATGTGGTCCCGATTTGGAAAA  111
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  GlnLeuLeuArgGluSerLeuAlaAlaGluAlaGluArgGlyCysGlyProArgPheGlyLys  20

Qy    112  CACTTGCTGTCATATTGCCCCATGCCCGAGAAGACATTCACCACCACCCAGGAGGGTGG  171
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21  HisLeuLeuSerTyrCysProMetProGluLysThrPheThrThrThrProGlyGlyTrp  40

Qy    172  CTG  174
           |||
Db     41  Leu  41
```

```
RESULT 11
US-09-201-227A-27
;   Sequence 27, Application US/09201227A
;   Patent No. 6468770
;   GENERAL INFORMATION:
;   APPLICANT: Keyes, Linda N.
;   APPLICANT: Doberstein, Stephen K.
;   APPLICANT: Buchman, Andrew R.
;   TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS OF D. MELANOGASTER INSULIN-LIKE
;   TITLE OF INVENTION: GENES AND USES THEREOF
;   FILE REFERENCE: 7326-066
;   CURRENT APPLICATION NUMBER: US/09/201,227A
;   CURRENT FILING DATE: 1998-11-30
```

(Top-left rotated fragment)

```
Qy    112  CACTTGCTGTCATATTGCCCCATGCCCGAGAAGACATTCACCACCACCCAGGAGGGTGG  171
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21  HisLeuLeuSerTyrCysProMetProGluLysThrPheThrThrThrProGlyGlyTrp  40

Qy    172  CTG  174
           |||
Db     41  Leu  41
```

Db |||||||||||||||||||||||||||||||
          21 AspAspGlyThrSerValLysLeuCysThr 30

RESULT 13
US-09-599-564A-12
; Sequence 12, Application US/09599564A
; Patent No. 6362318
; GENERAL INFORMATION:
;   APPLICANT: KOMAN, Ahment
;   APPLICANT: CHASSIN, Dorine
;   APPLICANT: BELLET, Dominique
;   TITLE OF INVENTION: NEW PROTEIN CALLED EPIL/PLACENTIN, PROCESS FOR THE
;   TITLE OF INVENTION: PREPARATION OF THIS PROTEIN AND PHARMACEUTICAL
;   TITLE OF INVENTION: COMPOSITION CONTAINING SUCH, DNA CODING FOR SAID
;   TITLE OF INVENTION: PROTEIN
;   FILE REFERENCE: 017753-127
;   CURRENT APPLICATION NUMBER: US/09/599,564A
;   CURRENT FILING DATE: 2000-06-23
;   PRIOR APPLICATION NUMBER: 09/174,465
;   PRIOR FILING DATE: 1998-10-19
;   PRIOR APPLICATION NUMBER: US 08/482,842
;   PRIOR FILING DATE: 1995-06-07
;   NUMBER OF SEQ ID NOS: 16
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
;   LENGTH: 30
;   TYPE: PRT
;   ORGANISM: Unknown
;   FEATURE:
;   OTHER INFORMATION: Description of Unknown Organism:EPIL - Early
;   OTHER INFORMATION: Placenta Insulin-Like Peptide
US-09-599-564A-12

ligment Scores:
red. No.:   3.35e-13   Length: 30
core: 171.00   Matches: 30
ercent Similarity: 100.00%   Conservative: 0
est Local Similarity: 100.00%   Mismatches: 0
uer% Match: 22.09%   Indels: 0
B: 4   Gaps: 0

S-09-518-842-1 (1-420) x US-09-599-564A-12 (1-30)

Qy   328 TCCCGCAAAAGAGAAGTGGACGTCACAGATTTGATCCATTCTGTTGTGAAGTAATTTGT 387
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 SerArgLysArgSerGlyArgHisArgPheAspProPheCysCysGluValIleCys 20

Qy   388 GACGATGGAACTTCAGTTAAATTATGTACA 417
          |||||||||||||||||||||||||||||||
Db   21 AspAspGlyThrSerValLysLeuCysThr 30

RESULT 14
S-09-201-227A-28
; Sequence 28, Application US/09201227A
; Patent No. 6468770
; GENERAL INFORMATION:
;   APPLICANT: Keyes, Linda N.
;   APPLICANT: Doberstein, Stephen K.
;   APPLICANT: Buchman, Andrew R.
;   TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS OF D. MELANOGASTER INSULIN-LIKE
;   TITLE OF INVENTION: GENES AND USES THEREOF
;   FILE REFERENCE: 7326-066
;   CURRENT APPLICATION NUMBER: US/09/201,227A
;   CURRENT FILING DATE: 1998-11-30
;   NUMBER OF SEQ ID NOS: 45
;   SOFTWARE: PatentIn version 3.0
; SEQ ID NO 28
;   LENGTH: 25
;   TYPE: PRT
;   ORGANISM: Homo sapiens
S-09-201-227A-28

ligment Scores:

Pred. No.:   3.15e-10   Length: 25
Score: 147.00   Matches: 25
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 18.99%   Indels: 0
DB: 4   Gaps: 0

US-09-518-842-1 (1-420) x US-09-201-227A-28 (1-25)

Qy   343 AGTGGACGTCACAGATTGATCCATTCTGTTGTGAAGTAATTTGTGACGATGGAACTTCA 402
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 SerGlyArgHisArgPheAspProPheCysCysGluValIleCysAspAspGlyThrSer 20

Qy   403 GTTAAATTATGTACA 417
          |||||||||||||||
Db   21 ValLysLeuCysThr 25

RESULT 15
US-08-950-720A-12
; Sequence 12, Application US/08950720A
; Patent No. 6046028
; GENERAL INFORMATION:
;   APPLICANT: Conklin, Darrell C.
;   APPLICANT: Lofton-Day, Catherine E.
;   APPLICANT: Lok, Si
;   APPLICANT: Jaspers, Stephen R.
;   TITLE OF INVENTION: INSULIN HOMOLOG
;   NUMBER OF SEQUENCES: 17
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: ZymoGenetics, Inc.
;   STREET: 1201 Eastlake Avenue East
;   CITY: Seattle
;   STATE: WA
;   COUNTRY: USA
;   ZIP: 98102
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/950,720A
;   FILING DATE:
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Sawislak, Deborah A
;   REGISTRATION NUMBER: 37,438
;   REFERENCE/DOCKET NUMBER: 96-09
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 206-442-6672
;   TELEFAX: 206-442-6678
;   TELEX:
; INFORMATION FOR SEQ ID NO: 12:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 185 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: No. 6046028e
US-08-950-720A-12

Alignment Scores:
Pred. No.:   7.33e-07   Length: 185
Score: 123.50   Matches: 53
Percent Similarity: 35.35%   Conservative: 17
Best Local Similarity: 26.77%   Mismatches: 55
Query Match: 15.96%   Indels: 73
DB: 3   Gaps: 7

US-09-518-842-1 (1-420) x US-08-950-720A-12 (1-185)

```
)y    1  ATGGCCAGCCTGTTCCGGTCCTATCTGCCAGCAATCTGCTGCTGCTGAGCCAACTCCTT  60
              |||||    |||||              :::   ||||||||||:::|||
)b    1  MetProArgLeuPhePheHisLeuLeuGlyyValCysLeuLeuLeuAsnGlnPheSer  20

)y   61  --------------------------------------------AGAGAAAGCCTA  72
                                                         |||||    :::
)b   21  ArgAlaValAlaAspSerTrpMetGluGluValIleLysLeuCysGlyArgGluLeuVal  40

)y   73  GCAGCAGAGCTGAGGGGATCTGGT--CCCGATTTGGAAAACACTTGCTGTCATATTGC  129
              |||:::::  |||||         :::   |||    ||||||
)b   41  ArgAlaGlnIleAlaIleLeuCysGlyMetSerThrTrpSerLysSerLeuSerGlnGlu  60

)y  130  CCCATGCCTGAGAAGACATTCACCACACCCCAGGAGGTGGCTGCTGGAATCTGGACGT  189
              |||:::                  |||||
)b   61  AspAlaProGln--------------ThrPro---------------------Arg  67

)y  190  CCC---AAAGAAATGGTGTCAACTCCAACAACAAAGATGGACAAAGCCTTAGGTACGACA  246
              ||||||||||:::||      |||||||||    :::
)b   68  ProValAlaGluIleValProSerPheIleAspLysAspThrArgIleAsnMetMet  87

)y  247  TCAGAATTCATTCCTAATTGTCTACCAGAGCTGAAGAAACCACTGTCTCGAAGGGCAGCCA  306
              ||||||||||  |||||              ||||||||||||
)b   88  SerGluPheValAlaAsnLeuProGlnLeuLeuLysLeuThrLeuSerGluMetGlnPro  107

)y  307  TCATTG-----------------------------------------------------  312
              :::|||
)b  108  AlaLeuProGlnLeuGlnHisValProValLeuLysAspSerSerLeuLeuPheGlu  127

)y  313  ------AAGAAATAATACTT--------------------------------------  327
              ||||||    :::|||
)b  128  GluPheLysLeuLeuIleArgAsnArgGlnSerGlnAlaAlaAspSerSerProSerGlu  147

)y  328  ---------------------TCCCGCAAAAGAGAAGTGGACGTCACAGATTT  360
                                   |||||||||||||||
)b  148  LeuLysTyrLeuLeuAspThrHisSerArgLysArgLysGlnLeuTyrSerAlaLeu  167

)y  361  GATCCATTCTGTTGTGAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGT  414
              |||||| |||    |||               :::    |||
)b  168  AlaAsnLysCysCysHisValGlyCysThrIleArgSerLeuArgAlaArgPheCys  185
```

earch completed: October 9, 2003, 12:20:32
ob time : 23 secs

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2003  Compugen Ltd.

M nucleic - protein search, using frame_plus_n2p model

un on:     October  9, 2003, 12:09:01 ; Search time 16.5 Seconds
                                 (without alignments)
                                 2394.087 Million cell updates/sec

itle:        US-09-518-842-1
Perfect score: 774
equence:      1 ATGGCCAGCCTGTTCCGGTC..........CAGTTAAATTATGTACATAG 420

coring table: BLOSUM62
              Xgapop   10.0 , Xgapext  0.5
              Xgapop   10.0 , Xgapext  0.5
              Fgapop    6.0 , Fgapext  7.0
              Delop     6.0 , Delext   7.0

earched:     127863 seqs, 47026705 residues

otal number of hits satisfying chosen parameters:       255726

atabase :     SwissProt_41:*

1inimum DB seq length: 0
1aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%
                Maximum Match 100%
                Listing first 45 summaries

ommand line parameters:
MODEL=frame+n2p.model.-DEV=xlp
Q=/cgn2_1/USPTO_spool/US09518842/runat_09102003_111035_13047//app_query.fasta_1.583
DB=SwissProt_41:-QFMT=fastan -SUFFIX=rsp -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
USER=US09518842 @CGN_1_1_30 @runat_09102003_111035_13047 -NCPU=3 -ICPU=3
NO_MMAP -LARGEQUERY-NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=30 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                       SUMMARIES
```

esult No.	Score	Query Match %	Length	DB	ID	Description
1	739	95.5	139	1	INL4_HUMAN	Q14641 homo sapien
2	180	23.3	182	1	RELX_PIG	P01348 sus scrofa
3	151	19.5	185	1	REL1_MOUSE	P47932 mus musculu
4	147.5	19.1	182	1	RELX_HORSE	P22969 equus caball
5	123.5	16.0	185	1	REL2_HUMAN	P04090 homo sapien
6	122.5	15.8	185	1	RELX_MACMU	P19884 macaca mula
7	120	15.5	186	1	RELX_RAT	P01347 rattus norv
8	119.5	15.4	185	1	REL1_HUMAN	P04808 homo sapien
9	113.5	14.7	166	1	REL2_PANTR	P51455 pan troglod
10	111.5	14.4	178	1	RELH_RABIT	P51456 oryctolagus
11	108.5	14.0	166	1	REL1_PANTR	P51454 pan troglod
12	104.5	13.5	177	1	RELX_CANFA	Q9tm8 canis famil
13	95.5	12.3	177	1	RELX_MESAU	Q64171 mesocricetu
14	80	10.3	240	1	IBP6_HUMAN	P24592 homo sapien
15	80	10.3	447	1	NUOF_BUCAP	Q8Ky3 buchnera ap
16	78.5	10.1	890	1	LPN1_HUMAN	Q14693 homo sapien
17	76.5	9.9	160	1	RELX_CAVPO	P51453 cavia porce
18	75	9.7	224	1	XP4_XENLA	Q00223 xenopus lae

```
c 19    75   9.7    836  1  DPOL_HPBDU  P03162 duck hepati
c 20    75  10.1    993  1  RPN1_YEAST  P38764 saccharomyc
c 21    75  10.1   1711  1  PTPO_RAT    Q64612 rattus norv
c 22    74.5 10.0    564  1  FEP1_SCHPO  Q10134 schizosacch
        74.5  9.6   1531  1  PMPD_CHLTR  O84818 chlamydia t
c 24    74    9.9    857  1  AR56_CANAL  P78586 c_arg5.6 pr
  25    73.5  9.5    273  1  NADE_STAAM  Q99sx5 staphylococ
c 26    73.5  9.9    296  1  PERX_BRARA  P00434 brassica ra
  27    73    9.4   1395  1  SP41_YEAST  P38904 saccharomyc
c 28    73    9.8   1744  1  CO4_HUMAN   P01028 homo sapien
  29    72.5  9.4    292  1  YHFR_SALTI  Q8z4m8 salmonella
  30    72.5  9.4    467  1  IL6A_PIG    O18796 sus scrofa
c 31    72    9.7    403  1  METM_CAEEL  O17680 caenorhabdi
  32    72    9.3    752  1  YUU3_SCHPO  Q9urt2 schizosacch
  33    72    9.3   3358  1  PGCV_MOUSE  Q62059 mus musculu
  34    71.5  9.2    292  1  YFHR_SALTY  Q8zn39 salmonella
  35    71.5  9.6   1021  1  PSKR_DAUCA  Q81pb4 daucus caro
  36    71    9.2    653  1  ACSA_SYNY3  Q55404 synechocyst
  37    71    9.2    957  1  UVRA_BACSU  O34863 bacillus su
c 38    70.5  9.1    271  1  VG18_BPP22  P03687 bacteriopha
  39    70.5  9.5   1704  1  VILD_DICDI  Q8wq85 dictyosteli
c 40    70    9.0    283  1  PANC_SCHPO  Q09673 schizosacch
  41    70    9.0    344  1  KLP3_MOUSE  Q60980 mus musculu
c 42    70    9.4    388  1  METK_ACACA  Q95032 acanthamoeb
c 43    70    9.4    486  1  CDSN_HUMAN  Q15517 homo sapien
c 44    70    9.4    670  1  YFGG_SCHPO  O13854 schizosacch
c 45    70    9.4   1356  1  ROM2_YEAST  P51862 saccharomyc
```

```
                                          ALIGNMENTS

RESULT 1
INL4_HUMAN
ID   INL4_HUMAN      STANDARD;      PRT;      139 AA.
AC   Q14641;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Early placenta insulin-like peptide precursor (EPIL) (Placentin)
DE   (Insulin-like peptide 4).
GN   INSL4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=96115599; PubMed=8666396;
RA   Chassin D., Laurent A., Janneau J.-L., Berger R., Bellet D.;
RT   "Cloning of a new member of the insulin gene superfamily (INSL4)
RT   expressed in human placenta.";
RL   Genomics 29:465-470(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
```

RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP PROCESSING.
RX MEDLINE=97430657; PubMed=9284764;
RA Bellet D., Lavaissiere L., Mock P., Laurent A., Sabourin J.C.,
RA Bedossa P., Le Bouteiller P., Frydman R., Troalen F., Bidart J.M.;
RT "Identification of pro-EPIL and EPIL peptides translated from
RT insulin-like 4 (INSL4) mRNA in human placenta.";
RL J. Clin. Endocrinol. Metab. 82:3169-3172(1997).
RN [4]
RP TISSUE SPECIFICITY.
RX MEDLINE=98411035; PubMed=9740319;
RA Laurent A., Rouillac C., Delezoide A.L., Giovangrandi Y., Vekemans M.,
RA Bellet D., Abitbol M., Vidaud M.;
RT "Insulin-like 4 (INSL4) gene expression in human embryonic and
RT trophoblastic tissues.";
RL Mol. Reprod. Dev. 51:123-129(1998).
CC -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN TROPHOBLAST DEVELOPMENT
CC   AND IN THE REGULATION OF BONE FORMATION.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN PLACENTA, UTERUS AND IN FETAL
CC   PERICHONDRIUM.
CC -!- DEVELOPMENTAL STAGE: HIGHLY EXPRESSED IN THE EARLY PLACENTA.
CC   EXPRESSION OF EPIL PEPTIDES IN THE VILLOUS CYTOTROPHOBLAST IS
CC   DIFFERENT FROM THAT DISPLAYED BY THE SYNCYTIOTROPHOBLAST. IN FETAL
CC   TISSUES IT WAS IDENTIFIED IN THE PERICHONDRIUM OF ALL FOUR LIMBS,
CC   VERTEBRAE, AND RIBS. IT WAS ABUNDANT IN INTERBONE LIGAMENTS.
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; L34838; AAB08516.1; -.
DR EMBL; BC026254; AAH26254.1; -.
DR Genew; HGNC:6087; INSL4.
DR MIM; 600910; -.
DR GO; GO:0005615; C:extracellular space; TAS.
DR GO; GO:0005625; C:soluble fraction; TAS.
DR GO; GO:0005159; F:insulin-like growth factor receptor binding. . .; TAS.
DR GO; GO:0008283; P:cell proliferation; TAS.
DR GO; GO:0007267; P:cell-cell signaling; TAS.
DR GO; GO:0007275; P:development; TAS.
DR GO; GO:0007565; P:pregnancy; TAS.
DR GO; GO:0007165; P:signal transduction; TAS.
DR InterPro; IPR004825; Ins/IGF/relax.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Signal.
FT SIGNAL   1   17   POTENTIAL.
FT CHAIN   18   139   EARLY PLACENTA INSULIN-LIKE PEPTIDE.
FT CHAIN   18   58   EARLY PLACENTA INSULIN-LIKE PEPTIDE B
FT   CHAIN.
FT PROPEP   59   109   C PEPTIDE.
FT CHAIN   110   139   EARLY PLACENTA INSULIN-LIKE PEPTIDE A
FT   CHAIN.
FT DISULFID   31   125   INTERCHAIN (BY SIMILARITY).
FT DISULFID   43   138   INTERCHAIN (BY SIMILARITY).
FT DISULFID   124   129   BY SIMILARITY.
SQ SEQUENCE   139 AA; 15445 MW; 47FB61F6F86C1342 CRC64;

Alignment Scores:
Pred. No.:   [5]
Score: 9.72e-71
Score: 739.00   Length: 139
Percent Similarity: 100.00%   Matches: 139
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 95.48%   Mismatches: 0
DB: Indels: 0
                                               Gaps: 0

US-09-518-842-1 (1-420) x INL4_HUMAN (1-139)

```
Qy      1 ATGGCCAGCCTGTCCGGTCCTATCTGCCAGCAATCTGGCTGCTGCTGAGCCAACTCCTT   60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MetAlaSerLeuPheArgSerTyrLeuProAlaIleTrpLeuLeuLeuSerGlnLeuLeu
```
```
Qy     61 AGAGAAAGCCTAGCAGCAGAGCTGAGGGGATGTGGTCTCCCGATTCGGAAAACACTTGCTG  120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 ArgGluSerLeuAlaAlaGluLeuArgGlyCysGlyProArgPheGlyLysHisLeuLeu   40
```
```
Qy    121 TCATATTGCCCATGCCTGAGAAGACATTCACCACCACCCCAGGAGGTGGCTGCTGGAA     180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     41 SerTyrCysProMetProGluLysThrPheThrThrThrProGlyGlyTrpLeuLeuGlu   60
```
```
Qy    181 TCTGGACGTCCAAAGAAATGGTGTCAACCTCAACACAACAAAGATGACAAGCCTTAGT    240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 SerGlyArgProLysGluMetValSerThrSerAsnAsnLysAspGlyGlnAlaLeuGly   80
```
```
Qy    241 ACGACATCAGAATTCATTCCTAATTGTCACCAGAGCTGAGAAACCACTGTCTGAAGGG    300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81 ThrThrSerGluPheIleProAsnLeuSerProGluLeuLysProLeuSerGluGly    100
```
```
Qy    301 CAGCCATCATTGAAGAAAAATAATACTTTCCCGCAAAAAGACGAAGTGACGTCACAGATTT  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    101 GlnProSerLeuLysIleLeuLeuSerArgLysLysArgSerGlyArgHisArgPhe    120
```
```
Qy    361 GATCCATTCTGTTGTGAAGTAATTGTGACGATGGAACTTCAGTTAAATATGTACA    417
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AspProPheCysCysGluValIleCysAspAspGlyThrSerValLysLeuCysThr  139
```

RESULT 2
RELX_PIG
ID RELX_PIG   STANDARD;   PRT;   182 AA.
AC P01348;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Prorelaxin precursor.
GN RLN.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87308187; PubMed=2442155;
RA Haley J., Crawford R., Hudson P., Scanlon D., Tregear G., Shine J.;
RA Niall H.;
RT "Porcine relaxin. Gene structure and expression.";
RL J. Biol. Chem. 262:11940-11946(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=83157118; PubMed=6897721;
RA Haley J., Hudson P., Scanlon D., John M., Cronk M., Shine J.;
RA Tregear G., Niall H.;
RT "Porcine relaxin: molecular cloning and cDNA structure.";
RL DNA 1:155-162(1982).
RN [3]
RP PRELIMINARY SEQUENCE OF 25-54 AND 161-182.
RX MEDLINE=77213067; PubMed=876374;
RA James R., Niall H., Kwok S., Bryant-Greenwood G.;
RT "Primary structure of porcine relaxin: homology with insulin and
RT related growth factors.";
RL Nature 267:544-546(1977).
RN [4]
RP SEQUENCE OF 25-51.
RX MEDLINE=77157271; PubMed=851452;
RA Schwabe C., McDonald J.K., Steinetz B.G.;
RT "Primary structure of the B-chain of porcine relaxin.";
RL Biochem. Biophys. Res. Commun. 75:503-510(1977).
RN [5]
RP SEQUENCE OF 25.

```
RX   MEDLINE=77134136; PubMed=843375;
RA   Schwabe C., McDonald J.K.;
RT   "Demonstration of a pyroglutamyl residue at the N terminus of the B-
RT   chain of porcine relaxin.";
RL   Biochem. Biophys. Res. Commun. 74:1501-1504(1977).
RN   [6]
RP   SEQUENCE OF 161-182.
RX   MEDLINE=76231539; PubMed=938497;
RA   Schwabe C., McDonald J.K., Steinetz B.G.;
RT   "Primary structure of the A chain of porcine relaxin.";
RL   Biochem. Biophys. Res. Commun. 70:397-405(1976).
RN   [7]
RP   DISULFIDE BONDS.
RX   MEDLINE=77235040; PubMed=887933;
RA   Schwabe C., McDonald J.K.;
RT   "Relaxin: a disulfide homolog of insulin.";
RL   Science 197:914-915(1977).
RN   [8]
RP   3D-STRUCTURE MODELING.
RX   MEDLINE=78092399; PubMed=622170;
RA   Isaacs N.W., James R., Niall H., Bryant-Greenwood G., Dodson G.G.,
RA   Evans A., North A.C.T.;
RT   "Relaxin and its structural relationship to insulin.";
RL   Nature 271:278-281(1978).
CC   -!- FUNCTION: RELAXIN IS AN OVARIAN HORMONE THAT ACTS WITH ESTROGEN
CC       TO PRODUCE DILATATION OF THE BIRTH CANAL IN MANY MAMMALS.
CC   -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC       DISULFIDE BONDS.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; K01088; AAA31114.1; -.
DR   EMBL; J02792; AAA31115.1; -.
DR   EMBL; A16593; CAA01295.1; -.
DR   EMBL; A06852; CAA00600.1; -.
DR   PIR; A90934; RXPG.
DR   PDB; 1RLX; 15-OCT-94.
DR   PDB; 2RLX; 15-OCT-94.
DR   PDB; 3RLX; 15-OCT-94.
DR   PDB; 4RLX; 15-OCT-94.
DR   InterPro; IPR004825; Ins/IGF/relax.
DR   Pfam; PF00049; Insulin; 1.
DR   SMART; SM00078; IIGF; 1.
DR   PROSITE; PS00262; INSULIN; 1.
KW   Insulin family; Hormone; Signal; 3D-structure;
KW   Pyrrolidone carboxylic acid.
FT   SIGNAL        1     24
FT   CHAIN        25     56       RELAXIN B CHAIN.
FT   PROPEP       57    154       CONNECTING PEPTIDE.
FT   CHAIN       161    182       RELAXIN A CHAIN.
FT   MOD_RES      25     25       PYRROLIDONE CARBOXYLIC ACID.
FT   DISULFID     34    169       INTERCHAIN.
FT   DISULFID     46    182       INTERCHAIN.
FT   DISULFID    168    173
FT   CONFLICT     47     47       G -> GVWS (IN REF. 4).
FT   CONFLICT     51     54       WGRT -> TWGR (IN REF. 3).
FT   CONFLICT    116    116       S -> L (IN REF. 1).
FT   CONFLICT    170    170       Q -> E (IN REF. 6).
FT   STRAND       30     31
FT   TURN         35     35
FT   HELIX        36     46
FT   HELIX        47     49
FT   HELIX       164    168
FT   TURN        169    170
FT   STRAND      173    174
FT   HELIX       175    178
FT   HELIX       179    182
SQ   SEQUENCE   182 AA;   20818 MW;  20736EB089F13AB4 CRC64;
```

```
Alignment Scores:
Pred. No.:                               178
Score:                        180.00     182
Percent Similarity:    38.02%    3.74e-11
Best Local Similarity: 31.77%
Query Match:           23.26%
DB:                        1

Length:       182
Matches:       61
Conservative:  12
Mismatches:    55
Indels:        64
Gaps:           6
```

```
US-09-518-842-1 (1-420) x RELX_PIG (1-182)

Qy    1 ATGGCCAGCCGTTCCGGTCCTATCTGCCAGCAATCTGGCTGCTGCTGAGCCAACTCCTT 60
          |||  |||||||||||||||||        ::|||||||||||||        |||||||||
Db    1 MetProArgLeuPhe---SerTyrLeuLeuGlyValTrpLeuLeuSerGlnLeuPro 19

Qy   61 AGAGAA----------------AGCCTAGCAGCAGAGCTGAGGGGATGTGGTCCCGATTGGA 108
          ||||||                   |||          :::::: |||||
Db   20 ArgGluIleProGlyGlnSerThrAsnAspPheIleLysAlaCysGlyArgGluLeuVal 39

Qy  109 AAACACTTGCTGTCATATTGCCCCATGCCTGAGAAGACATTCACCACCCACCAGGAGGG 168
          :::                              |||
Db   40 ArgLeuTrpValGluIleCysGlySerVal---------------------Ser 50

Qy  169 TGG------------------------------CTGCTGGAATCTGCGACGTCCCAAAGAA 198
          |||                               |||:::||| |||  |||
Db   51 TrpGlyArgThrAlaLeuSerLeuGluGluProGlnLeuGlnThrGlyProProAlaGlu 70

Qy  199 ATGGTGTCAACCTCCAACAACAAAGATGGACAGCCTTAGGTACGACATCAGAATTCATT 258
          |||     ::|||        |||||  |||
Db   71 ThrMetProSerSerIleThrLysAspAlaGluIleLeuLysMetLeuGluPheVal 90

Qy  259 CCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCATTG----- 312
          |||||||||||||||||      |||||||||  ||||||||||||
Db   91 ProAsnLeuProGlnGluLeuLysAlaThrLeuSerGluArgGlnProSerLeuArgGlu 110

Qy  313 ----------------------------------------AAGAAAATAATA 324
                                                   ||||||||||||
Db  111 LeuGlnGlnSerAlaSerLysAspSerAsnLeuAsnPheGluGluLysPheLysLysIleIle 130

Qy  325 CTT------------------------------------ 327
          |||
Db  131 LeuAsnArgGlnAsnGluAlaGluAspLysSerLeuLeuGluLeuLysAsnLeuGlyLeu 150

Qy  328 ---------TCCCGCAAAAGAAGAAGTGGACGTCACAGATTTGATCCATTCGTTGTGAA 378
                   |||||||||||||||        |||
Db  151 AspLysHisSerArgLysArgLeuPheArgMetThrLeuSerGluLysCysCysGln 170

Qy  379 GTAATTGTGACGATGGAACTTCAGTTAAATTATGT 414
          |||        ::|||||
Db  171 ValGlyCysIleArgLysAspIleAlaArgLeuCys 182
```

```
RESULT 3
REL1_MOUSE
ID   REL1_MOUSE     STANDARD;      PRT;    185 AA.
AC   P47932;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Prorelaxin 1 precursor.
GN   RLN1 OR RLN OR RLX.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Swiss; TISSUE=Ovary;
RX   MEDLINE=93199663; PubMed=8452637;
RA   Evans B.A., John M., Fowler K.J., Summers R.J., Cronk M.,
RA   Shine J., Tregear G.W.;
```

```
Db   141 AspGlySerProProGlyLeuLysTyrLeuGlnSerAspThrHisSerArgLysLysArg 160
         ||||||||||||||||||||
Qy   343 AGTGGACGTCACAGATTGATCCATTCTGTTGTGAAGTAATTGTGACGATGGAACTTCA 402
         ||||||                                   |||
Db   161 GluSerGlyGlyLeuMetSerGlnGlnCysCysHisValGlyCysSerArgArgSerIle 180
         ||||||                                   :::
Qy   403 GTTAAATTA 411
         |||||||
Db   181 AlaLysLeu 183
```

RESULT 4
RELX_HORSE

```
ID   RELX_HORSE     STANDARD;      PRT;   182 AA.
AC   P22969; Q28907;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Prorelaxin precursor (RXN).
GN   RLN.
OS   Equus caballus (Horse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX   NCBI_TaxID=9796;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Hokkaido; TISSUE=Placenta;
RA   Min K., Shiota K., Ogawa T.;
RT   "Molecular cloning of equine preprorelaxin cDNA.";
RL   J. Reprod. Dev. 42:171-178(1996).
RN   [2]
RP   SEQUENCE OF 32-174 FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=95319320; PubMed=7543295;
RA   Klonisch T., Ryan P.L., Yamaghiro S., Porter D.G.;
RT   "Partial complementary deoxyribonucleic acid cloning of equine relaxin
RT   messenger ribonucleic acid, and its localization within the equine
RT   placenta.";
RL   Biol. Reprod. 52:1307-1315(1995).
RN   [3]
RP   SEQUENCE OF 26-53 AND 163-182.
RC   TISSUE=Placenta;
RX   MEDLINE=91275796; PubMed=2055195;
RA   Stewart D.R., Nevins B., Hadas E., Vandlen R.;
RT   "Affinity purification and sequence determination of equine relaxin.";
RL   Endocrinology 129:375-383(1991).
CC   -!- FUNCTION: RELAXIN IS AN OVARIAN HORMONE THAT ACTS WITH ESTROGEN
CC       TO PRODUCE DILATATION OF THE BIRTH CANAL IN MANY MAMMALS.
CC   -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC       DISULFIDE BONDS.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC   ----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; AB002201; BAA19069.1; -.
DR   EMBL; S78800; AAB35036.1; -.
DR   HSSP; P01348; 1RLX.
DR   InterPro; IPR004825; Ins/IGF/relax.
DR   Pfam; PF00049; Insulin; 1.
DR   SMART; SM00078; IlGF; 1.
DR   PROSITE; PS00262; INSULIN; 1.
KW   Insulin family; Hormone; Signal.
FT   SIGNAL        1     25
FT   CHAIN        26     53       RELAXIN B CHAIN.
FT   PROPEP       54    156       CONNECTING PEPTIDE.
FT   CHAIN       161    182       RELAXIN A CHAIN.
```

```
RT   "The mouse relaxin gene: nucleotide sequence and expression.";
RL   J. Mol. Endocrinol. 10:15-23(1993).
CC   -!- FUNCTION: RELAXIN IS AN OVARIAN HORMONE THAT ACTS WITH ESTROGEN
CC       TO PRODUCE DILATATION OF THE BIRTH CANAL IN MANY MAMMALS.
CC   -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC       DISULFIDE BONDS.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC   ----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; Z27088; CAA81611.1; -.
DR   PIR; S48082; S48082.
DR   MGD; MGI:97931; Rln1.
DR   InterPro; IPR004825; Ins/IGF/relax.
DR   Pfam; PF00049; Insulin; 1.
DR   SMART; SM00078; IlGF; 1.
DR   PROSITE; PS00262; INSULIN; FALSE_NEG.
KW   Insulin family; Hormone; Signal.
FT   SIGNAL        1     22       BY SIMILARITY.
FT   CHAIN        23     57       RELAXIN B CHAIN (PROBABLE).
FT   PROPEP       58    156       CONNECTING PEPTIDE (PROBABLE).
FT   CHAIN       161    185       RELAXIN A CHAIN (PROBABLE).
FT   DISULFID     36    171       INTERCHAIN (BY SIMILARITY).
FT   DISULFID     48    185       INTERCHAIN (BY SIMILARITY).
FT - DISULFID    170    175       BY SIMILARITY.
SQ   SEQUENCE   185 AA;  20570 MW;  2B9B753EBB85087B CRC64;
```

```
Alignment Scores:
Pred. No.:                      4.62e-08      Length:         185
Score:               151.00                   Matches:         50
Percent Similarity:   37.70%                  Conservative:    19
Best Local Similarity: 27.03%                 Mismatches:      68
Query Match:          19.51%                  Indels:          46
DB:                      1                    Gaps:             3
```

US-09-518-842-1 (1-420) x RELL_MOUSE (1-185)

```
Qy    1 ATGGCCAGCCGTCGTTCCGGTCCCATCGCCAGCAATCTGGCTGCTGCTGAGCCAACTCCTT 60
        ||||::||     ||                                 |||||||||||||
Db    1 MetSerSerArgPheLeuLeuGlnLeuLeuGlyLeuPheTrpLeuLeuSerGlnProCys 20

Qy   61 AGAGAAAGCCTAGCAGCAGAG--------------CTGAGGGGATGTGTGTCCCCGATTT 105
        |||   :::::               :::||||   ||||:|
Db   21 ArgThrArgValSerGluGluTrpMetAspGlyPheIleArgMetCysGlyArgGluTyr 40

Qy  106 GGAAAACACTTGCTGTCATATTGCCCCATGCTGAGAAGACATTCACCACCACCCCAGGA 165
        :::                                     :::
Db   41 AlaArgGluLeuIleLysIleCysGlySerValGlyArgLeuAlaLeuSerGlnGlu 60

Qy  166 GGGTGGCTGCTGGAACTTGGACGTCCCAAAGAAATGGTGTCAACCTCCAACAACAAAGAT 225
        |||   ::: :::                  ||||||||
Db   61 GluProAlaLeuLeuAlaArgGlnAlaThrGlnAlaValProSerPheIleAsnLysAsp 80

Qy  226 GGACAAGCCTTAGGTACGACCATCAGAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAA 285
        :::   |||                     ::||||||||
Db   81 AlaGluProPheAspThrThrLeuLysCysCysHisAsnLeuSerGluGluLeuLysAla 100

Qy  286 CCACTGTCGAGGGGCAGCCATCATTGAAGAAATA--------ATACTTT------ 327
        ||||||||||   ||||||    |||     ::::      :::||||||
Db  101 ValLeuSerGluAlaGlnAlaSerLeuProGluLeuGlnHisAlaProValLeuSerAsp 120

Qy  327 ------------------------------------- 327
Db  121 SerValValSerLeuGluGlyPheLysLysThrLeuHisAspArgLeuGlyGluAlaGlu 140

Qy  328 -----------------TCCCCGCAAAAGAGAGA 342
```

```
T   DISULFID      35    169    INTERCHAIN (BY SIMILARITY).
T   DISULFID      47    182    INTERCHAIN (BY SIMILARITY).
T   DISULFID     168    173    BY SIMILARITY.
T   CONFLICT      66     66    A -> V (IN REF. 2).
T   CONFLICT     133    133    L -> Q (IN REF. 2).
Q   SEQUENCE     182 AA;  20721 MW;  E5C9414303A838B8 CRC64;

lignment Scores:
red. No.:              1.09e-07     Length:          182
core:                  147.50       Matches:          43
ercent Similarity:     52.59%       Conservative:     18
est Local Similarity:  37.07%       Mismatches:       44
Query Match:           19.06%       Indels:           11
DB:                                 Gaps:              4

US-09-518-842-1 (1-420) x RELX_HORSE (1-182)

y    1 ATGGCCAGCCGCTGTTCGGCTCCTATCTGCCGCAATCTGGCTGCTGCTGAGCCAACTCCTT 60
        |||    ||||||||    ||||||    ::::::    ||||||||||||||||||||
b    1 MetArgArgLeuPheLeuSerHisValLeuGlyAlaAlaTrpLeuLeuSerGlnLeuPro 20

y   61 AGAGAAAGCCTAGCAGCAGAG----------CTGAGGGGATGTGGTCCCCGATTTGA 108
        ||||||                 ::        :::::    ||||||
b   21 ArgGluLeuSerGlyGlnLysProAspAspValIleLysAlaCysGlyArgGluLeuAla 40

y  109 AAACACTTGCTGTCATATTGCCCCATGCCT-----GAGAAGACATTC------ACCACC 156
        :::      :::                           |||
b   41 ArgLeuArgIleGluIleCysGlySerLeuSerTrpLysThrValLeuArgLeuGlu 60

y  157 ACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCAAAGAAATGGTGTCAACCTCCAAC 216
        ||||||          ||||||||:::||||:::||||||:::::||||
b   61 GluProGly--------LeuGluAlaGlyVinProValGluValSerSerSerIle 77

y  217 AACAAGATGGACGAAGCCTTAGGTACGACATCAGAATTCATTCCTAATTGTCACCAGAG 276
        :::|||||||    :::||||||    |||         |||    ||
b   78 SerLysAspAlaGluLeuAlaLeuAsnThrLysLeuGlyLeuAsnSerAsnLeuProLysGlu 97

y  277 CTGAAGAAACACCTGTCTGAAGGGCAGCCATCATTGAAGAAAATAATA 324
        ||||||          ||||||||          :::::::::
b   98 GlnLysAlaThrLeuSerLeuArgGlnProSerTrpArgGluLeuLeu 113
```

```
ESULT 5
EL2_HUMAN
D   REL2_HUMAN     STANDARD;     PRT;     185 AA.
C   P04090; Q99936; Q9UCX3; Q9UOJ2;
C   P04090; 01-NOV-1986 (Rel. 03, Created)
C   01-NOV-1986 (Rel. 03, Last sequence update)
C   15-SEP-2003 (Rel. 42, Last annotation update)
E   Prorelaxin H2 precursor.
E   RLN2.
S   Homo sapiens (Human).
C   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
X   NCBI_TaxID=9606;
N   [1]
P   SEQUENCE FROM N.A. (ISOFORM 1), AND TISSUE SPECIFICITY.
X   MEDLINE=85051298; PubMed=6548702;
A   Hudson P., John M., Crawford R., Haralambidis J., Scanlon D.,
A   Gorman J., Tregear G., Shine J., Niall H.;
T   "Relaxin gene expression in human ovaries and the predicted structure
T   of a human preprorelaxin by analysis of cDNA clones.";
L   EMBO J. 3:2333-2339(1984).
N   [2]
P   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
A   Sehra H.;
L   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
N   [3]
P   PARTIAL SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND TISSUE
P   SPECIFICITY.
C   TISSUE=Prostate;
X   MEDLINE=96328899; PubMed=8735594;
A   Gunnersen J.M., Fu P., Roche P.J., Tregear G.W.;
T   "Expression of human relaxin genes: characterization of a novel
```

```
RT   alternatively-spliced human relaxin mRNA species.";
RL   Mol. Cell. Endocrinol. 118:85-94(1996).
RN   [4]
RP   SEQUENCE OF 1-15 FROM N.A.
RA   Garibay-Tupas J.;
RT   "Characterization of the human relaxin H1 and H2 5'-flanking
RT   regions.";
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   PARTIAL SEQUENCE OF 25-51.
RC   TISSUE=Semen;
RX   MEDLINE=92241162; PubMed=1572287;
RA   Winslow J.W., Shih A., Bourell J.H., Weiss G., Reed B., Stults J.T.,
RA   Goldsmith L.T.;
RT   "Human seminal relaxin is a product of the same gene as human luteal
RT   relaxin.";
RL   Endocrinology 130:2660-2668(1992).
RN   [6]
RP   SEQUENCE OF 25-53 AND 162-185.
RX   MEDLINE=91167739; PubMed=2076464;
RA   Winslow J.W., Griffin P.R., Rinderknecht E., Vandlen R.L.;
RT   "Structural characterization by mass spectrometry of native and
RT   recombinant human relaxin.";
RL   Biomed. Environ. Mass Spectrom. 19:655-664(1990).
RN   [7]
RP   SYNTHESIS.
RX   MEDLINE=91250367; PubMed=2040595;
RA   Buellesbach E.E., Schwabe C.;
RT   "Total synthesis of human relaxin and human relaxin derivatives by
RT   solid-phase peptide synthesis and site-directed chain combination.";
RL   J. Biol. Chem. 266:10754-10761(1991).
RN   [8]
RP   X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).
RX   MEDLINE=92015205; PubMed=1656049;
RA   Eigenbrot C., Randal M., Quan C., Burnier J., O'Connell L.,
RA   Rinderknecht E., Kossiakoff A.A.;
RT   "X-ray structure of human relaxin at 1.5 A. Comparison to insulin and
RT   implications for receptor binding determinants.";
RL   J. Mol. Biol. 221:15-21(1991).
CC   -!- FUNCTION: RELAXIN IS AN OVARIAN HORMONE THAT ACTS WITH ESTROGEN TO
CC       PRODUCE DILATATION OF THE BIRTH CANAL IN MANY MAMMALS. MAY BE
CC       INVOLVED IN REMODELING OF CONNECTIVE TISSUES DURING PREGNANCY,
CC       PROMOTING GROWTH OF PUBIC LIGAMENTS AND RIPENING OF THE CERVIX.
CC   -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC       DISULFIDE BONDS.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=P04090-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=P04090-2; Sequence=VSP_002711, VSP_002712;
CC   -!- TISSUE SPECIFICITY: Expressed in the ovary during pregnancy. Also
CC       expressed in placenta, decidua and prostate.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X00948; CAA25460.1; -.
DR   EMBL; AL135786; CAC04177.1; -.
DR   EMBL; AL135786; CAC04176.1; -.
DR   EMBL; S83200; AAD14429.1; -.
DR   EMBL; AF104935; AAD21961.1; -.
DR   EMBL; A17315; CAA01324.1; -.
DR   EMBL; A06925; CAA00602.1; -.
DR   PIR; A05092; A60982.
DR   PDB; 6RLX; 31-OCT-93.
```

```
DR   Genew; HGNC:10027; RLN2.
DR   MIM; 179740; -.
DR   GO; GO:0007565; P:pregnancy; TAS.
DR   InterPro; IPR004825; Ins/IGF/relax.
DR   Pfam; PF00049; Insulin; 1.
DR   SMART; SM00078; IGF; 1.
DR   PROSITE; PS00262; INSULIN; 1.
KW   Insulin family; Hormone; Multigene family; Signal; 3D-structure;
KW   Alternative splicing; Pyrrolidone carboxylic acid.
FT   SIGNAL        1     24
FT   CHAIN        25     53       RELAXIN B CHAIN.
FT   PROPEP       56    157       CONNECTING PEPTIDE.
FT   CHAIN       162    185       RELAXIN A CHAIN.
FT   MOD_RES     162    162       PYRROLIDONE CARBOXYLIC ACID.
FT   DISULFID     35    172       INTERCHAIN.
FT   DISULFID     47    185       INTERCHAIN.
FT   DISULFID    171    176       INTERCHAIN.
FT   VARSPLIC     71    117       EIVPSFINKDTETINMMSEFVANLPQELKLTLSEMQPALPQ
FT                                LQQHVP -> GDFIOTVSLGISPDGGKALRTGSCFTREFLG
FT                                ALSKLCHPSSTKIQKP (in isoform 2).
FT                                /FTId=VSP_002711.
FT   VARSPLIC    118    185       Missing (in isoform 2).
FT                                /FTId=VSP_002712.
FT   HELIX        27     29
FT   STRAND       31     32
FT   HELIX        36     49
FT   HELIX       164    173
FT   TURN        174    174
FT   STRAND      176    177
FT   HELIX       178    182
FT   TURN        183    185
SQ   SEQUENCE    185 AA;  21042 MW;  AC73DBDE2090091B CRC64;

Alignment Scores:
Pred. No.:
Score:                    3.94e-05
Percent Similarity:       123.50                 Length:        185
Best Local Similarity:    35.35%                 Matches:        53
Query Match:              26.77%                 Conservative:   17
                          15.96%                 Mismatches:     55
DB:                                              Indels:         73
                                                 Gaps:            7

US-09-518-842-1 (1-420) x REL2_HUMAN (1-185)

Qy     1 ATGGCCAGCCTGTTCCGGTCCTATCTGCCAGCAATCTGGCTGCTGCTGAGCCAACTCCTT  60
           |||||
Db     1 MetProArgLeuPhePheHisLeuLeuGlyValCysLeuLeuLeuAsnGlnPheSer  20

Qy    61 ----------------------------------------------AGAGAAAGCCTA  72
                                                          |||||:::|||
Db    21 ArgAlaValAlaAlaAspSerTrpMetGluGluValIleLysLeuLysGlyArgGluLeuVal  40

Qy    73 GCAGCAGAGCTGAGGGGATGTGGT--CCCCGATTTGGAAAACACTTGCTGTCATATTGC 129
           |||||           ||||||    |||                ||| ||||
Db    41 ArgAlaGlnAlaIleAlaIleCysGlyMetSerThrTrpSerLysThrArgSerLeuSerGlnGlu  60

Qy   130 CCCATGCTGCTGGAAGACATTCACCACCACCCAGGAGGAGGTGGCTGCTGGAATCTGGACGT 189
           |||||                                      |||||
Db    61 AspAlaProGln----------------ThrPro----------------Arg  67

Qy   190 CCC--AAAGAAATGGTGTCAACCTCCAACAACAAAGATGACAAGACCCTAGGTACGACA 246
           |||   |||||:||      |||||||       ::: |||:|
Db    68 ProValAlaGluIleValProSerPheIleAsnLysAspThrGluThrIleAsnMetMet  87

Qy   247 TCAGAATTCATTCCTAATTGTCACCAGAGCTGAGAAGAACCACTGTCTGAAGGGCAGCCA 306
           ||||:::                  ||||||     ||||||
Db    88 SerGluPheValAlaAsnLeuProGlnGluLeuLysLeuThrLeuSerGluMetGlnPro 107

Qy   307 TCATTG------------------------------------------------ 312
           :::||
Db   108 AlaLeuProGlnGlnHisValProValLeuLysAspSerLeuSerLeuPheGlu 127

Qy   313 -----AAGAAAATAATACTT------------------------------ 327
              ||||||:::|||
```

```
RESULT 6
RELX_MACMU
ID   RELX_MACMU    STANDARD;    PRT;    185 AA.
AC   P19884;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Prorelaxin precursor.
GN   RLN.
OS   Macaca mulatta (Rhesus macaque).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9544;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90073957; PubMed=2599381;
RA   Crawford R.J., Hammond V.E., Roche P.J., Johnston P.D., Tregear G.W.;
RT   "Structure of rhesus monkey relaxin predicted by analysis of the
RT   single-copy rhesus monkey relaxin gene.";
RL   J. Mol. Endocrinol. 3:169-174(1989).
CC   -!- FUNCTION: RELAXIN IS AN OVARIAN HORMONE THAT ACTS WITH ESTROGEN TO
CC       PRODUCE DILATATION OF THE BIRTH CANAL IN MANY MAMMALS. MAY BE
CC       INVOLVED IN REMODELING OF CONNECTIVE TISSUES DURING PREGNANCY,
CC       PROMOTING GROWTH OF PUBIC LIGAMENTS AND RIPENING OF THE CERVIX.
CC   -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC       DISULFIDE BONDS.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR   PIR; A34936; A34936.
DR   HSSP; P04090; 6RLX.
DR   InterPro; IPR004825; Ins/IGF/relax.
DR   Pfam; PF00049; Insulin; 1.
DR   SMART; SM00078; IGF; 1.
DR   PROSITE; PS00262; INSULIN; 1.
KW   Insulin family; Hormone; Signal.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23     53       RELAXIN B CHAIN (PROBABLE).
FT   PROPEP       56    157       CONNECTING PEPTIDE (PROBABLE).
FT   CHAIN       161    185       RELAXIN A CHAIN (PROBABLE).
FT   DISULFID     35    172       INTERCHAIN (BY SIMILARITY).
FT   DISULFID     47    185       INTERCHAIN (BY SIMILARITY).
FT   DISULFID    171    176       BY SIMILARITY.
SQ   SEQUENCE    185 AA;  20895 MW;  7B3C5D21B57E185C CRC64;

Alignment Scores:
Pred. No.:                1
Score:                    5.03e-05
Percent Similarity:       122.50                 Length:        185
Best Local Similarity:    34.34%                 Matches:        51
Query Match:              25.76%                 Conservative:   17
                          15.83%                 Mismatches:     57
DB:                                              Indels:         73
                                                 Gaps:            6

US-09-518-842-1 (1-420) x RELX_MACMU (1-185)

Qy     1 ATGGCCAGCCTGTTCCGGTCCTATCTGCCAGCAATCTGGCTGCTGCTGAGCCAACTCCTT  60
           |||                                        ::: |||||
Db     1 MetProArgLeuPheLeuPheHisLeuLeuGlyValCysLeuLeuLeuAsnGlnPheSer  20

Qy    61 -------------------------------------------AGAGAAAGCCTA  40
                                                      |||||
Db    21 ArgAlaValAlaAlaAlaLysTrpMetAspArgValIleLeuLysAlaCysGlyArgGluLeuVal  40

Db   128 GluPheLysLysLeuIleArgArgAsnArgGlnSerGluAlaAlaAlaAspSerSerProSerGlu 147
Db   328 ----------------------------------------TCCCGCAAAAAGAGAAGTGGACGTCACAGATTT 360
                                                   ||||||
Qy
Db   148 LeuLysTyrLeuGlyLeuAspThrHisSerArgLysLysArgGlnLeuTyrSerAlaLeu 167
           |||||
Qy   361 GATCCATTCTGTTGTGAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGT 414
           |||||                ::: |||||
Db   168 AlaAsnLysCysCysHisValGlyCysThrLysSerLeuAlaArgPheCys 185
```

```
CC      ----------------------------------------------------------------
DR      EMBL; J00780; AAA42029.1; -.
DR      EMBL; V01264; CAA24578.1; -.
DR      PIR; A01614; RXRT.
DR      HSSP; P01348; 1RLX.
DR      InterPro; IPR004825; Ins/IGF/relax.
DR      Pfam; PF00049; Insulin; 1.
DR      SMART; SM00078; IIGF; 1.
DR      PROSITE; PS00262; INSULIN; 1.
KW      Insulin family; Hormone; Signal; Pyrrolidone carboxylic acid.
FT      SIGNAL        1     22
FT      CHAIN        23     57       RELAXIN B CHAIN.
FT      PROPEP       58    158       CONNECTING PEPTIDE.
FT      CHAIN       163    186       RELAXIN A CHAIN.
FT      MOD_RES     163    163       PYRROLIDONE CARBOXYLIC ACID.
FT      DISULFID     36    173       INTERCHAIN (BY SIMILARITY).
FT      DISULFID     48    186       INTERCHAIN (BY SIMILARITY).
FT      DISULFID    172    177       BY SIMILARITY.
SQ      SEQUENCE    186 AA;  20489 MW;  08BAEC79BCF0880F CRC64;

Alignment Scores:
Pred. No.:              9.3e-05
Score:                120.00
Percent Similarity:    33.16%          Length:        186
Best Local Similarity: 23.83%          Matches:        46
Query Match:           15.50%          Conservative:   18
                                       Mismatches:     67
DB:                                    Indels:         62
                                       Gaps:            5

US-09-518-842-1 (1-420) x RELX_RAT (1-186)

Qy     1 ATGGCCAGCCTGTTCCGGTCCTATCTGCCAGCAATCTGGCTGCTGCTGAGCCAACTCCTT    60
         ||:::||||                    |||    |||||| | ||||||
Db     1 MetSerSerArgLeuLeuLeuLeuGlnLeuLeuGlyPheTrpLeuSerGlnProCys    20

Qy    61 AGAGAAGCCTAGCAGCAGAG------------------CTGAGGGGATGTGGTCCCCGATTT   105
         |||::|||    ::|::|||                   |||:|||  |||||||||
Db    21 ArgAlaAArgValSerGluTrpMetGlnValIleGlnValCysGlyArgGlyTyr    40

Qy   106 GGAAAACACTTGCTGTCATATTGCCCCATGCTGGAGAAGACATTCACCACCACCACCAGGA   165
         ::                ::  :::                     |||
Db    41 AlaArgAlaTrpIleGluValCys------------------GlyAlaSerValGly    53

Qy   166 GGGTGGCTGCTGGAATCTGACGTCCC---------------------AAAGAAATG    201
         |||||||||    |||   |||                        |||:::
Db    54 ArgLeuAlaLeuSerGlnGluGluGlnLeuProAlaProLeuAlaArgGlnAlaThrAlaGluVal    73

Qy   202 GTGTCAACCTCCAACAACAAAGATGACAAGCCTTAGGTACGACATCAGAATTCATTCCT   261
         ||    |||       ::  :::    :::         |||   ||  |||||||
Db    74 ValProSerPheIleAsnLysAspAlaGluProPheAspMetThrLeuLysCysLeuPro    93

Qy   262 AATTTGTCACCAGAGCTGAAGAAACCACCTGTCTGAAGGGCAG-------------   303
         ||||||||    |||  ||  |||||||||||||:::
Db    94 AsnLeuSerGluGluGlyArgLysAlaAlaAlaLeuSerGlyLyArgAlaProPheProGluLeu   113

Qy   303 --------------------------------------------------------   303
Db   114 GlnGlnHisAlaProAlaLeuSerArgAspSerValSerLeuGluGlyPheLysLysThr   133

Qy   304 ----------------------------------------CCATCATTGAAGAAAATA---   321
                                                 ||| |||||||
Db   134 PheHisAsnGlnLeuGlyGluAlaGluAlaAspGlyGlyProGluLeuLysTyrLeuGly   153

Qy   322 -----ATACTTCCCGCAAAAGAAGAGAAGGTGACGTCACAGATTTGATCCATTCTGTTGT   375
              ||||||||||||:::
Db   154 SerAspAlaGlnSerArgLysLysArgGlnSerGlyAlaLeuLeuSerGluGlnCysCys   173

Qy   376 GAAGTAATTGTGACGATGAACTTCAGTTAAAATTATGT   414
         :::  ||||    :::  ::
Db   174 HisIleCysThrArgArgSerIleAlaLysLeuCys   186

RESULT 8
REL1_HUMAN
ID      REL1_HUMAN               STANDARD;      PRT;         185 AA.
```

(second column, left portion — continuation of RESULT 7, ELX_RAT)

```
Qy    73 GCAGCAGAGCTGAGGGGATGTGGT---CCCGATTTGGAAAACACTTGCTGTCATATTGC   129
         ||:::::||| |||||                ||||||       ||||||
Db    41 ArgAlaGlnIleAlaIleCysGlyLysGlyLeuAsnGlnGlu    60

Qy   130 CCCATGCTGAGAGACATTCACCACCACCACCCCCAGGAGGGTGGCTGCTGGAATCTGACGT   189
         ||| |||          ||||                              ----Arg    67
Db    61 AspAlaProLeuLysPro

Qy   190 CCC---AAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACACA   246
         |||               ||:  :::     |||  :::
Db    68 ProAlaAlaGlulleValProSerLeulleAsnGlnAspThrGlulleAspAsnMetMet    87

Qy   247 TCAGAATTCATTCCTAATTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCA   306
         ||||||||::: |||||||                ||||| |||||
Db    88 SerGluPheValAlaAsnLeuProGlnGluLeuLysThrLeuSerGluGluArgGlnPro   107

Qy   307 TCATTG-----------------------------------------   312
         :::||
Db   108 AlaLeuSerGluLeuGlnGlnHisValProValLeuLysAspAsnLeuSerPheGlu   127

Qy   313 -----AAGAAAATAATACTT-----------------------   327
         ||||||||||||
Db   128 GluPheLysLysIleIleGluArgLysArgGlnSerGluAlaThrAspSerProSerGlu   147

Qy   328 -------------------------TCCCGCAAAAGAGAAGTGGACGTCACAGATTT   360
         ||:::::||||||
Db   148 LeuArgSerLeuGlyLeuAspThrHisSerArgArgLysArgGlnLeuTyrMetThrLeu   167

Qy   361 GATCCATTCTGTTGTGAAGTAATTGTGACGATGGAACTTCAGTTAAAATTATGT   414
         |||||| ||| |||
Db   168 SerAsnLysCysHisIleIleGlyCysThrLysLysSerLeuAlaLysPheCys   185
```

```
RESULT 7
ELX_RAT
ID      RELX_RAT               STANDARD;      PRT;         186 AA.
```

```
AC   P04808; Q99936; Q9UQJ1;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Prorelaxin H1 precursor.
GN   RLN1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RX   MEDLINE=85051298; PubMed=6548702;
RA   Hudson P., John M., Crawford R., Haralambidis J., Scanlon D.,
RA   Gorman J., Tregear G., Shine J., Niall H.;
RT   "Relaxin gene expression in human ovaries and the predicted structure
RT   of a human preprorelaxin by analysis of cDNA clones.";
RL   EMBO J. 3:2333-2339(1984).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RX   MEDLINE=83141755; PubMed=6298628;
RA   Hudson P., Haley J., John M., Cronk M., Crawford R., Haralambidis J.,
RA   Tregear G., Shine J., Niall H.;
RT   "Structure of a genomic clone encoding biologically active human
RT   relaxin.";
RL   Nature 301:628-631(1983).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RA   Sehra H.;
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Prostate;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [5]
RP   PARTIAL SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND TISSUE
RP   SPECIFICITY.
RC   TISSUE=Prostate;
RX   MEDLINE=96328899; PubMed=8735594;
RA   Gunnersen J.M., Fu P., Roche P.J., Tregear G.W.;
RT   "Expression of human relaxin genes: characterization of a novel
RT   alternatively-spliced human relaxin mRNA species.";
RL   Mol. Cell. Endocrinol. 118:85-94(1996).
RN   [6]
RP   SEQUENCE OF 1-11 FROM N.A.
RA   Garibay-Tupas J.;
RT   "Characterization of the human H1 relaxin 5'-flanking region.";
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: RELAXIN IS AN OVARIAN HORMONE THAT ACTS WITH ESTROGEN TO
CC       PRODUCE DILATATION OF THE BIRTH CANAL IN MANY MAMMALS. MAY BE
CC       INVOLVED IN REMODELING OF CONNECTIVE TISSUES DURING PREGNANCY,
CC       PROMOTING GROWTH OF PUBIC LIGAMENTS AND RIPENING OF THE CERVIX.
CC   -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC       DISULFIDE BONDS.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=P04808-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=P04808-2; Sequence=VSP_002709, VSP_002710;
CC   -!- TISSUE SPECIFICITY: Prostate. Not expressed in placenta, decidua
CC       or ovary.
CC   -!- MISCELLANEOUS: H1 RELAXIN MAY BE A PSEUDOGENE.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X00949; CAA25461.1; -.
DR   EMBL; V00578; CAA23839.1; -.
DR   EMBL; V00577; CAA23838.1; -.
DR   EMBL; AF104934; AAD21967.1; -.
DR   EMBL; A06926; CAA00603.1; -.
DR   EMBL; A06846; CAA00599.1; -.
DR   EMBL; A07364; CAA00658.1; -.
DR   EMBL; A17329; CAA01325.1; -.
DR   EMBL; AL135786; CAC04179.1; -.
DR   EMBL; BC005956; AAH05956.1; -.
DR   EMBL; S83200; AAD14429.1; -.
DR   PIR; B05092; A44559.
DR   HSSP; P04090; 6RLX.
DR   Genew; HGNC:10026; RLN1.
DR   MIM; 179730; -.
DR   GO; GO:0005180; F:peptide hormone; TAS.
DR   InterPro; IPR004825; Ins/IGF/relax.
DR   Pfam; PF00049; Insulin; 1.
DR   SMART; SM00078; IIGF; 1.
DR   PROSITE; PS00262; INSULIN; 1.
KW   Insulin family; Hormone; Multigene family; Signal;
KW   Alternative splicing; Polymorphism.
FT   SIGNAL        1     22
FT   CHAIN        23     53       RELAXIN B CHAIN (PROBABLE).
FT   PROPEP       56    158       CONNECTING PEPTIDE (PROBABLE).
FT   CHAIN       163    185       RELAXIN A CHAIN (PROBABLE).
FT   DISULFID     35     47       INTERCHAIN (BY SIMILARITY).
FT   DISULFID    171    185       INTERCHAIN (BY SIMILARITY).
FT   DISULFID     47    176       INTERCHAIN (BY SIMILARITY).
FT   VARSPLIC     71    117       EIVPSFINKDTETIILMLEFIANLPPELKAALSEROPSLPE
FT                                LQQYVP -> GDFIQTVSLGISPDGGKALRTGSCFTREFLG
FT                                ALSKLYHPSSTKIQKL (in isoform 2).
FT                                /FTId=VSP_002709.
FT   VARSPLIC    118    185       Missing (In isoform 2).
FT                                /FTId=VSP_002710.
FT   VARIANT      28     28       K -> M (in dbSNP:618066).
FT                                /FTId=VAR_011962.
SQ   SEQUENCE   185 AA;  21145 MW;  B318628ABFEC7142 CRC64;
```

```
Alignment Scores:
Pred. No.:               0.000105      Length:              185
Score:                 119.50         Matches:              52
Percent Similarity:     37.11%        Conservative:         20
Best Local Similarity:  26.80%        Mismatches:           57
Query Match:            15.44%        Indels:               65
DB:                         1         Gaps:                  7

US-09-518-842-1 (1-420) x REL1_HUMAN (1-185)

Qy      1 ATGGCCAGCCTGTTCCGGTCCTATCTGCCAGCAATCTGGCTGCTGCTGAGCCAACTCCTT  60
          |||      ||||||||                |::||||      ||||||||::::|||
Db      1 MetProArgLeuPheLeuPheHisLeuLeuGluPheCysLeuLeuLeuAsnGlnPheSer  20
```

```
Qy    61 AGAGAAAGCCTAGCAGCAGAG------------CTGAGGGGATGTGGTCCCCGATTT 105
         |||  :::::::|||||:::                  ||||||
Db    21 Arg---AlaValAlaLysTrpAspAspValIleLysLeuCysGlyArgGluLeu 39

Qy   106 GGAAAACACTTGCTTGCTGTCATATTGCCCCATG-----------------CCTGAG 141
         :::|||                              |||
Db    40 ValArgAlaGlnIleAlaIleCysGlyLysSerThrTrpSerLysSerGln 59

Qy   142 AAGACATTCACCACACCCCAGGAGGTGGCTGCTGGGAATCTGACGTCCC--AAAGAA 198
         :::                                |||:::        |||
Db    60 GluAspAlaProGlnThrPro-----------------ArgProValAlaGlu 71

Qy   199 ATGTGTCAACCTCCAACAACAAAGATGACAAGCCTTAGGTACGACATCAGAATTCATT 258
         :::||                  :::  |||  :::    |||||||||
Db    72 IleValProSerPheIleAsnLysAspThrGluThrIleIleMetMetSerGlyPheIle 91

Qy   259 CCTAATTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCATG------ 312
         |||:::|||  |||||||    |||  ||| |||||||:::|||
Db    92 AlaAsnLeuProGlnLeuLysLeuAlaLeuSerGluArgGlnProSerLeuProGlu 111

Qy   313 -------------------------------------------------AAGAAA 318
                                                           ||||||
Db   112 LeuGlnGlnTyrValProAlaLeuLysAspSerSerAsnLeuSerPheGluGluLysLys 131

Qy   319 ATATTACTTTCCCGC------------------------------------ 333
         ::|||   ||  |||
Db   132 LeuIleArgAsnArgGlnSerGluAlaAlaAlaAspSerAsnProSerGluLeuLysTyrLeu 151

Qy   334 -------------------------AAAAGAGAAGTGGACGTCACAGATTTGATCCATTCTGT 372
                                  ||||||||||           ||
Db   152 GlyLeuAspThrHisSerGlnLysLysLysArgArgProTyrValAlaLeuPheGluLysCys 171

Qy   373 TGTGAAGTAATTGTGACGATGGAACTTCAGTTAAATTATGT 414
         ::|||   :::        |||        ||  |||
Db   172 CysLeuIleGlyCysThrLysLysSerLeuAlaLysTyrCys 185
```

RESULT 9
EL2_PANTR

```
ID   REL2_PANTR     STANDARD;     PRT;    166 AA.
AC   P51455; P79267;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Prorelaxin H2 precursor (Fragment).
GN   RNL2 OR RLX2.
OS   Pan troglodytes (Chimpanzee).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX   NCBI_TaxID=9598;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1), AND TISSUE SPECIFICITY.
RC   TISSUE=Placenta;
RX   MEDLINE=94238260; PubMed=8182365;
RA   Evans B.A., Fu P., Tregear G.W.;
RT   "Characterization of two relaxin genes in the chimpanzee.";
RL   J. Endocrinol. 140:385-392(1994).
RN   [2]
RP   PARTIAL SEQUENCE FROM N.A. (ISOFORM 2), AND TISSUE SPECIFICITY.
RC   TISSUE=Ovary, and Placenta;
RX   MEDLINE=96328899; PubMed=8735594;
RA   Gunnersen J.M., Fu P., Roche P.J., Tregear G.W.;
RT   "Expression of human relaxin genes: characterization of a novel
RT   alternatively-spliced human relaxin mRNA species.";
RL   Mol. Cell. Endocrinol. 118:85-94(1996).
CC   -!- FUNCTION: RELAXIN IS AN OVARIAN HORMONE THAT ACTS WITH ESTROGEN TO
CC       PRODUCE DILATATION OF THE BIRTH CANAL IN MANY MAMMALS. MAY BE
CC       INVOLVED IN REMODELING OF CONNECTIVE TISSUES DURING PREGNANCY,
CC       PROMOTING GROWTH OF PUBIC LIGAMENTS AND RIPENING OF THE CERVIX.
CC   -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC       DISULFIDE BONDS.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=P51455-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=P51455-2; Sequence=VSP_002713, VSP_002714;
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN THE CORPUS LUTEUM OF PREGNANCY
CC       AND IN THE PLACENTA.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z27245; CAA81758.1; -.
DR   EMBL; S83209; AAD14430.1; -.
DR   PIR; S42786; S42786.
DR   HSSP; P04090; 6RLX.
DR   InterPro; IPR004825; Ins/IGF/relax.
DR   Pfam; PF00049; Insulin; 1.
DR   SMART; SM00078; IlGF; 1.
DR   PROSITE; PS00262; INSULIN; 1.
KW   Insulin family; Hormone; Multigene family; Signal; Polymorphism;
KW   Alternative splicing.
FT   NON_TER       1      1
FT   SIGNAL       <1      5
FT   CHAIN         6     34       RELAXIN B CHAIN (PROBABLE).
FT   PROPEP       37    138       CONNECTING PEPTIDE (PROBABLE).
FT   CHAIN       143    166       RELAXIN A CHAIN (PROBABLE).
FT   DISULFID     16    153       INTERCHAIN (BY SIMILARITY).
FT   DISULFID     28    166       INTERCHAIN (BY SIMILARITY).
FT   DISULFID    152    157       BY SIMILARITY.
FT   VARSPLIC     52     84       EIVPSFINKDTETINMWSEFVANLPOBLKLTLS -> DFIQ
FT                                TVSLGISPDGGKALRTGSCFTREFLGALS (in isoform
FT                                2).
FT                                /FTId=VSP_002713.
FT   VARSPLIC     85    166       Missing (in isoform 2).
FT                                /FTId=VSP_002714.
FT   VARIANT      30     34       KSTWS -> MSTLG (IN SOME ALLELES).
SQ   SEQUENCE   166 AA;  18760 MW;  22085lE3134CEBDB CRC64;

Alignment Scores:
Pred. No.:               0.000449      Length:             166
Score:                 113.50          Matches:             45
Percent Similarity:     36.42%         Conservative:        14
Best Local Similarity:  27.78%         Mismatches:          46
Query Match:            14.66%         Indels:              57
DB:                                    Gaps:                 6

US-09-518-842-1 (1-420) x REL2_PANTR (1-166)

Qy    61 AGAGAAAGCCTAGCAGCAGAGCTGCTGGGGATGTGGT---CCCGATTGGAAAAACACTTG 117
         ||||||             ::  :::|||||||||||       |||
Db    18 ArgGluLeuValArgAlaAlaGlnIleAlaIleCysGlyLysSerThrTrpSerLysArgSer 37

Qy   118 CTGTCATATTGCCCCATGCCTGAGAAGACATTCACCACACCCCAGGAGGTGGCTGCTG 177
         |||||                              |||:::
Db    38 LeuSerGlnGluAspAlaProGln-----------------ThrPro----------- 47

Qy   178 GAATCTGACGTCCC---AAAGAAATGTGTCAACCTCCAACAACAAAGATGACAAGCC 234
         |||||              :::|||        ::::|||||||||
Db    48 ---------ArgProValAlaGluIleValProSerPheIleAsnLysAspThrGluThr 64

Qy   235 TTAGGTACGACATCAGAATTCATTCCTAATTGTCACCAGAGCTGAAGAAACCACTGTCT 294
         :::    |||||||||:::|||     :::|||:||  |||||||    |||
Db    65 IleAsnMetMetSerGlyPheValAlaAsnLeuProGlnLeuLysLeuThrLeuSer 84

Qy   295 GAAGGGCAGCCATCATG------------------------------ 312
         ||| |||||||:::|||
Db    85 GluMetGlnProAlaLeuProGlnLeuGlnLeuGlnTyrValProValLeuLysAspSerSer 104
```

```
Pred. No.:              0.000743        Length:        178
Score:                  111.50          Matches:       49
Percent Similarity:     32.77%          Conservative:  9
Best Local Similarity:  27.68%          Mismatches:    60
Query Match:            14.41%          Indels:        59
DB:                     1               Gaps:          6

US-09-518-842-1 (1-420) x RELH_RABIT (1-178)

Qy     4 GCCAGCCTGCGTTCCGGTCCTATCGCCAGCAATCTGCCTGCTGCTGAGCCAA---------- 54
         ||||| ||||||  |||| ||||||  ||||  ||||||||||||| ||
Db     3 AlaLeuLeuPhe------------TyrLeuLeuGlyPheCysLeuLeuGlnValThrGly 20

Qy    55 ---CTCCTTAGAGAAAGCCTAGCAGCAGAGCTGAGGGGATGGTCCCCGATTTGGAAAA 111
            :::  ::::  ::::  :: ::::::::: ::: ||| :::  ::: |
Db    21 ArgValThrTyrGluTrpMetGluSerArgLysLysArgAsnAspPheValArg 40

Qy   112 CACTTGCTGTCATATTGC----------------------------CCCATGCTGAGAAG 144
         :: ||  |||| |||||||
Db    41 ThrAlaIleGluValCysGlyHisValHisLeuGluArgGluSerProSerProGluAsn 60

Qy   145 ACATTCACCACCACCCCAGGAGGGTGGCTGCTGCGGAATCGGACGTCCCAAGAAAATGGTG 204
         |||       ||| :: |:|       |||||||       |||  |||||
Db    61 ProPhe-------------------LeuSerGlyProAlaAlaGluThrVal 72

Qy   205 TCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACCATCAGAATTCATTCCTAAT 264
         :::|||| |||||| :::||| |:| ||| |||  ||||
Db    73 ProSerSerIleLysLysAspAlaGluAsnAlaAsnThrMetLeuGluSerIleProAsn 92

Qy   265 TTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCA------------ 309
         |||    ::||   :::|   ||  |::
Db    93 LeuProGlnGluLeuThrAlaThrLeuPheGluLysGlnProSerLysLeuTyrLeuGln 112

Qy   310 --------------------------------TTGAAGAAATAATA-- 324
                                         ||||||||||||
Db   113 TyrLeuProThrLeuLysLysSerAsnValSerPheGluPheLysLysIleIleGln 132

Qy   325 ----------------------------CTTTCCCGCAAA 336
                                     ||||||||||||
Db   133 AsnIleGlnArgValGlnGlySerSerAlaSerGluSerAsnThrPheSerArgLys 152

Qy   337 AAGAGAAGTGGACGTCACAGATTTGATCCATTCTGTTGTGAAGTAATTTGT 387
         ||||||        :::|||||||||||  |||::|::|||||
Db   153 LysArgGlnPheSerGluSerLeuProGluGluCysCysLysTyrGlyCys 169
```

```
Qy   313 ----------------------------AAGAAAATAATACTT----------------- 327
                                     ||||||:::||
Db   105 LeuLeuPheGluGluLysLysLeuIleArgAsnArgGlnSerGluAlaAlaAspser 124

Qy   328 ---------------------TCCCGCAAAAGAGAAGTGA 348
                              ||||||||||||||||
Db   125 SerProSerGluLeuLysTyrLeuGlyLeuAspThrHisSerArgLysLysArgGlnLeu 144

Qy   349 CGTCACAGATTTGATCCATTCTGTTGTGAAGTAATTGTGACGATGGAACTTCAGTTAAA 408
         ||||||         ||| ||| :::       ||: :::
Db   145 TyrSerAlaLeuAlaAsnLysCysCysHisValGlyCysThrLysTyrSerLeuAlaArg 164

Qy   409 TTATGT 414
         |||
Db   165 PheCys 166
```

```
RESULT 10
RELH_RABIT
ID   RELH_RABIT    STANDARD;      PRT;    178 AA.
AC   P51456;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Relaxin-like protein SQ10 precursor.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RC   TISSUE=Tracheobronchial epithelium;
RX   MEDLINE=93002619; PubMed=1333318;
RA   Jetten A.M., Bernacki S.H., Floyd E.E., Saunders N.A., Pieniazek J.,
RA   Lotan R.;
RT   "Expression of a preprorelaxin-like gene during squamous
RT   differentiation of rabbit tracheobronchial epithelial cells and its
RT   suppression by retinoic acid.";
RL   Cell Growth Differ. 3:549-556(1992).
CC   -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC       DISULFIDE BONDS (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Secreted (BY SIMILARITY).
CC   -!- INDUCTION: DURING SQUAMOUS CELL DIFFERENTIATION. REPRESSED BY
CC       RETINOIC ACID.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC  --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC  --------------------------------------------------------------------------
DR   EMBL; S45940; AAB23648.1; -.
DR   PIR; A49014; A49014.
DR   InterPro; IPR004825; Ins/IGF/relax.
DR   Pfam; PF00049; Insulin; 1.
DR   SMART; SM00078; IIGF; 1.
DR   PROSITE; PS00262; INSULIN; 1.
KW   Insulin family; Hormone; Signal.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21     52       RELAXIN-LIKE PROTEIN SQ10 B CHAIN
FT                                (POTENTIAL).
FT   PROPEP       53    150       CONNECTING PEPTIDE (POTENTIAL).
FT   CHAIN       155    178       RELAXIN-LIKE PROTEIN SQ10 A CHAIN
FT                                (POTENTIAL).
FT   DISULFID     34    165       INTERCHAIN (BY SIMILARITY).
FT   DISULFID     46    178       INTERCHAIN (BY SIMILARITY).
FT   DISULFID    164    169       BY SIMILARITY.
SQ   SEQUENCE   178 AA;  20294 MW;  F6A54D98A6B53211 CRC64;
```

```
RESULT 11
REL1_PANTR
ID   REL1_PANTR    STANDARD;      PRT;    166 AA.
AC   P51454;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Prorelaxin H1 precursor (Fragment).
GN   RNL1 OR RLX1.
OS   Pan troglodytes (Chimpanzee).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX   NCBI_TaxID=9598;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=94238260; PubMed=8182365;
RA   Evans B.A., Fu P., Tregear G.W.;
RT   "Characterization of two relaxin genes in the chimpanzee.";
RL   J. Endocrinol. 140:385-392(1994).
CC   -!- FUNCTION: RELAXIN IS AN OVARIAN HORMONE THAT ACTS WITH ESTROGEN TO
CC       PRODUCE DILATATION OF THE BIRTH CANAL IN MANY MAMMALS. MAY BE
CC       INVOLVED IN REMODELING OF CONNECTIVE TISSUES DURING PREGNANCY,
CC       PROMOTING GROWTH OF PUBIC LIGAMENTS AND RIPENING OF THE CERVIX.
CC   -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC       DISULFIDE BONDS.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN THE CORPUS LUTEUM OF PREGNANCY
```

ligment Scores:

CC　　BUT NOT IN THE PLACENTA.
CC　-!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC
CC　This SWISS-PROT entry is copyright. It is produced through a collaboration
CC　between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC　the European Bioinformatics Institute. There are no restrictions on its
CC　use by non-profit institutions as long as its content is in no way
CC　modified and this statement is not removed. Usage by and for commercial
CC　entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC　or send an email to license@isb-sib.ch).
CC
DR　EMBL; Z72225; CAA81739.1; -.
DR　PIR; S42783; S42783.
DR　HSSP; P04090; 6RLX.
DR　InterPro; IPR004825; Ins/IGF/relax.
DR　Pfam; PF00049; Insulin; 1.
DR　SMART; SM00078; IlGF; 1.
DR　PROSITE; PS00262; INSULIN; 1.
GW　Insulin family; Hormone; Multigene family; Signal.
FT　NON_TER　　1　　1
FT　SIGNAL　　1　　5　　POTENTIAL.
FT　CHAIN　　6　　34　　RELAXIN B CHAIN (PROBABLE).
FT　PROPEP　　37　　139　　CONNECTING PEPTIDE (PROBABLE).
FT　CHAIN　　143　　166　　RELAXIN A CHAIN (PROBABLE).
FT　DISULFID　16　　153　　INTERCHAIN (BY SIMILARITY).
FT　DISULFID　28　　166　　INTERCHAIN (BY SIMILARITY).
FT　DISULFID　152　　157　　BY SIMILARITY.
SQ　SEQUENCE　166 AA;　18730 MW;　7F469B1FB9259F4F CRC64;

Alignment Scores:
Pred. No.:　　　　　　0.00153
Score:　　　　　　　　108.50　　　　Length:　　　166
Percent Similarity:　37.65%　　　Matches:　　43
Best Local Similarity: 26.54%　　Conservative: 18
Query Match:　　　　　14.02%　　　Mismatches: 44
DB:　　　　　　　　　　　　　　　　　　Indels:　　57
　　　　　　　　　　　　　　　　　　　　Gaps:　　　7

US-09-518-842-1 (1-420) x REL1_PANTR (1-166)

y　　61 AGAGAAAGCTAGCAGCAGAGCTGAGGGATGTGGT--CCCCGATTTGGAAAACACTTG 117
　　　　　　|||||　　　:::::||||:::::　　　　　:::||　　　||||
b　　18 ArgGluLeuValArgAlaGlnIleAlaIleAlaIleCysGlyMetSerThrTrpSerLysArgSer 37

y　 118 CTGTCATATTGCCCCATGCTGAGRAGACATTCACCACCCCAGGAGGGTGCTGCTG 177
　　　　　||||||　　　:||::　　　　　　　|||||
b　　38 LeuSerGlnGlyAspAlaProGln---------------ThrPro------------- 47

y　 178 GAATCTGGACGTCCC--AAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCC 234
　　　　　||||||　　　||||:||||　　　:::　　||||||||　　　:::
b　　48 -------ArgProValAlaGluIleValProSerPheIleAsnLysAspThrGluThr 64

y　 235 TTAGGTACGACATCAGAATCATTCCTAATTGTCACCAGAGCTGAAGAAACCACTGTCT 294
　　　　　:::||　　　||||||　　　　||||　||||||||
b　　65 IleIleIleMetLeuGluPheIleAlaAsnLeuProProGluLeuLysAlaAlaLeuSer 84

y　 295 GAAGGGGCAGCCATCATTG--------------------- 312
　　　　　||||||||||||||
b　　85 GluArgGlnProSerLeuProGluProGlnGlnTyrValProAlaLeuLeuLysAspSerAsn 104

y　 313 --------------AAGAAAATAATAATACTTTCCCGCAAAAAGAGAAGT-------- 345
　　　　　　　　　　　　::||||||::　:::||||::
b　 105 LeuSerPheGluGluGluLysTyrLeuLysLeuIleArgArgAsnArgGlnSerGluAlaAlaAlaAspSer 124

y　 346 ---------------------------GGACGTCACAGATTTGATCCA 366
　　　　　　　　　　　　　　　　　:::　　||||
b　 125 AsnProSerGluLeuLysTyrLeuGlyLeuAspThrHisSerGlnLysLysArgGlnPro 144

y　 367 TTC---------------TGTTGTGAAGTAATTGTGACGATGGAACTTCAGTTAA 408
　　　　　:::　　　　　　　　||||
b　 145 TyrValAlaLeuPheGluLysCysCysLeuIleGlyCysThrLysArgSerLeuAlaAsn 164

y　 409 TTATGT 414
　　　　　|||

Db　　165 TyrCys 166

RESULT 12
RELX_CANFA
ID　RELX_CANFA　　　STANDARD;　　　PRT;　　177 AA.
AC　Q9TRM8; Q9N0Z7; Q9TRM9;
DT　16-OCT-2001 (Rel. 40, Created)
DT　16-OCT-2001 (Rel. 40, Last sequence update)
DT　16-OCT-2001 (Rel. 40, Last annotation update)
DE　Prorelaxin precursor.
GN　RLN.
OS　Canis familiaris (Dog).
OC　Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC　Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX　NCBI_TaxID=9615;
RN　[1]
RP　SEQUENCE FROM N.A.
RC　TISSUE=Placenta;
RX　MEDLINE=99150177; PubMed=10026098;
RA　Klonisch T., Hombach-Klonisch S., Froehlich C., Kauffold J.,
RA　Steger K., Steinetz B.G., Fischer B.;
RT　"Canine preprorelaxin: nucleic acid sequence and localization within
RT　the canine placenta.";
RL　Biol. Reprod. 60:551-557(1999).
RN　[2]
RP　SEQUENCE OF 26-60 AND 154-177.
RC　TISSUE=Placenta;
RX　MEDLINE=93000391; PubMed=1388669;
RA　Stewart D.R., Henzel W.J., Vandlen R.;
RT　"Purification and sequence determination of canine relaxin.";
RL　J. Protein Chem. 11:247-253(1992).
CC　-!- FUNCTION: Relaxin is an ovarian hormone that acts with estrogen
CC　　　to produce dilatation of the birth canal in many mammals.
CC　-!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC　　　DISULFIDE BONDS.
CC　-!- TISSUE SPECIFICITY: Placenta; syncytiotrophoblast.
CC　-!- SUBCELLULAR LOCATION: Secreted.
CC　-!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC
CC　This SWISS-PROT entry is copyright. It is produced through a collaboration
CC　between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC　the European Bioinformatics Institute. There are no restrictions on its
CC　use by non-profit institutions as long as its content is in no way
CC　modified and this statement is not removed. Usage by and for commercial
CC　entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC　or send an email to license@isb-sib.ch).
CC
DR　EMBL; AF233687; AAF60302.1; -.
DR　HSSP; P01348; 1RLX.
DR　InterPro; IPR004825; Ins/IGF/relax.
DR　Pfam; PF00049; Insulin; 1.
DR　SMART; SM00078; IlGF; 1.
DR　PROSITE; PS00262; INSULIN; 1.
KW　Insulin family; Hormone; Signal.
FT　SIGNAL　　1　　25
FT　CHAIN　　26　　60　　RELAXIN B CHAIN
FT　PROPEP　　63　　149　　CONNECTING PEPTIDE (BY SIMILARITY).
FT　CHAIN　　154　　177　　RELAXIN A CHAIN.
FT　DISULFID　34　　164　　INTERCHAIN.
FT　DISULFID　46　　177　　INTERCHAIN.
FT　CONFLICT　163　　168
FT　CONFLICT　49　　49　　I -> S (IN REF. 2).
SQ　SEQUENCE　177 AA;　20563 MW;　220B0EC99DD302A CRC64;

Alignment Scores:
Pred. No.:　　　　　　0.00413
Score:　　　　　　　　104.50　　　　Length:　　　177
Percent Similarity:　35.20%　　　Matches:　　43
Best Local Similarity: 24.02%　　Conservative: 20
Query Match:　　　　　13.50%　　　Mismatches: 55
DB:　　　　　　　　　　　　　　　　　　Indels:　　61
　　　　　　　　　　　　　　　　　　　　Gaps:　　　6

US-09-518-842-1 (1-420) x RELX_CANFA (1-177)

```
CC    or send an email to license@isb-sib.ch).
CC
DR    EMBL; S79879; AAB35555.1; -.
DR    InterPro; IPR004825; Ins/IGF/relax.
DR    Pfam; PF00049; Insulin; 1.
DR    SMART; SM00078; IlGF; 1.
KW    PROSITE; PS00262; INSULIN; 1.
KW    Insulin family; Hormone; Signal.
FT    SIGNAL        1     22       BY SIMILARITY.
FT    CHAIN        23     59       RELAXIN B CHAIN (PROBABLE).
FT    PROPEP       64    149       CONNECTING PEPTIDE (PROBABLE).
FT    CHAIN       154    177       RELAXIN A CHAIN (PROBABLE).
FT    DISULFID     36    168       INTERCHAIN (BY SIMILARITY).
FT    DISULFID     48    177       INTERCHAIN (BY SIMILARITY).
FT    DISULFID    163    168       BY SIMILARITY.
SQ    SEQUENCE    177 AA;  20007 MW;  692556 2BD8C66CCD CRC64;

Alignment Scores:
Pred. No.:                       177
Score:              0.0376        Matches:          33
Percent Similarity: 95.50         Conservative:     14
Best Local Similarity: 39.83%     Mismatches:       40
Query Match:         27.97%       Indels:           31
DB:                  12.34%       Gaps:              4
                     1

US-09-518-842-1 (1-420) x RELX_MESAU (1-177)

Qy      1 ATGGCCAGCCGTGTCCGGTCCTATCGCCAGCAATCGGCTGCTGTCGAGCCAACTCCTT 60
          ||||   ||      |||  |||:::||||||||||||||||||||||||||||||||
Db      1 MetLeuArgTrpPheLeuSerHisLeuGlyValTrpLeuLeuSerGlnLeuPro       20

Qy     61 AGAGAAAGCCTAGCAGCAGAG------CTGAGGGGATGTGGTCCCCGATTTCGAAAA   105
          ||||||||   ::                    |||||    ||  ::  ::
Db     21 ArgAlaArgValThrLysGluTrpLeuAspGluValIleHisValCysGlyArgGluTyr  40

Qy    106 GGAAAACACTTGCTGTCATATTGC-------------------------CCCATGCCT  138
          ::: |||||||||                                    |||:::
Db     41 ValArgAlaIleLeuAspIleCysAlaAlaThrValGlyLeuGluAlaProProLeuArg  60

Qy    139 GAGAAGACATTCACCACACCCAGGAGAGGGGTGGCTGCTGGAATCGGACGTCCCAAAGAA 198
          |||          |||     ||                          :::|||
Db     61 ArgArgArgMetThr----------------------------------GluGlu      67

Qy    199 ATGGTGTCAACCTCCAACAACAAACAATGACGAACAAGCCTTAGGTACGACCATCAGAATTCATT 258
          |||||||||||||||||||    :::     ::::: ||||||
Db     68 AlaValSerSerPheIleLysGluAspAlaGluProPheAspThr------------Met  83

Qy    259 CCTAATTTGTCACCAGACTGAAGAAACCACTGTCTGAAGGCAGCCAGCCATCATTG 312
          |||||||||||        |||    :::   ||||  |||||||||
Db     84 ProAsnLeuSerGluLysProLysThrAlaLeuProGluGlyHisProSerLeu 101
```

```
RESULT 14
IBP6_HUMAN
ID    IBP6_HUMAN     STANDARD;      PRT;    240 AA.
AC    P24592; Q14492;
DT    01-MAR-1992 (Rel. 21, Created)
DT    01-MAR-1992 (Rel. 21, Last sequence update)
DT    15-SEP-2003 (Rel. 42, Last annotation update)
DE    Insulin-like growth factor binding protein 6 precursor (IGFBP-6)
DE    (IBP-6) (IGF-binding protein 6).
GN    IGFBP6 OR IBP6.
OS    Homo sapiens (Human).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX    NCBI_TaxID=9606;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    TISSUE=Osteosarcoma;
RX    MEDLINE=91225006; PubMed=1709161;
RA    Kiefer M.C., Masiarz F.R., Bauer D.M., Zapf J.;
RT    "Identification and molecular cloning of two new 30-kDa insulin-like
RT    growth factor binding proteins isolated from adult human serum.";
RL    J. Biol. Chem. 266:9043-9049(1991).
```

```
Qy      1 ATGGCCAGCCGTGTCCGGTCCTATCGCCAGCAATCGGCTGCTGTCGAGCCAACTCCTT  60
          |||   |||      |||  |||  ||||||||||||||||||||||||||||||||
Db      1 MetLeuArgTrpPheLeuSerHisLeuGlyValTrpLeuLeuSerGlnLeuPro       20

Qy     61 AGAGAAAGCCTAGCAGCAGAG------CTGAGGGGATGTGGTCCCCGATTTCGAAAA  111
          ||||||||   :::                   |||||    ||  :::  :::
Db     21 ArgGlIleProAlaThrAspLysLysLeuLysAlaCysGlyArgAspTyrValArg     40

Qy    112 CACTTGCTGTCATATTGCCCCATGCCTGAGAAGACATTCACCACACCCAGGAGGG---  168
          :::                                  |||        :::
Db     41 LeuGlnIleGluValCys------------------------------GlySerIle    49

Qy    169 ---TGGCTGCTGGAATCGGACGTCCCAAA--------                       195
          |||          :::::: ||| ::  :::
Db     50 TrpTrpGlyArgLysAlaGlyGlnIleLeuArgGluArgArgGlnIleSerGluProLeuAla  69

Qy    196 GAAATGTGTCAACCTCCAACAACAAAGATGGACAAGCCCTAGGTACGACCATCAGAATTC  255
          |||||           :::              |||         ||
Db     70 GluValValProSerSerIleIleLysAspProGluIleLeuSerLeuMetLeuGlnSer   89

Qy    256 ATTCCTAATTTGTCACCAGACTGAGAGAAACCACTGTCTGAAGGCAGCCATCATTG---  312
          |||||             :::  ||||||||        |||:::   |||
Db     90 IleProGlyMetProGlnGluLeuArgIleAlaThrArgSerArgGlyLysLysLeuLeu  109

Qy    313 ------------------------------AAGAAAATA  321
                                         ||||||
Db    110 ArgGluLeuHisPheValLeuGluAspSerAsnLeuGluAsnLeuGluGluMetLysLysThr 129

Qy    322 ATACTTTCC----------------------------------  330
          |||:::
Db    130 PheLeuAsnThrGlnPheGluAlaGluAspLysSerLeuSerLeuAspLysLysHisPro  149

Qy    331 CGCAAAAAGAAGAAGTGACGTGCACAGATTTGATCCATTCTGTTGTGAAGTAATTTGT  387
          ||||||||||      |||||            |||
Db    150 ArgLysLysArgAspAsnTyrIleLysMetSerAspLysCysCysAsnValGlyCys 168
```

```
b
ESULT 13
ELX_MESAU
D     RELX_MESAU     STANDARD;      PRT;    177 AA.
C     Q64171;
X     MEDLINE=96115021; PubMed=7492700;
A     McCaslin R.B., Renegar R.H.;
T     "Determination of the prorelaxin nucleotide sequence and expression
T     of prorelaxin messenger ribonucleic acid in the golden hamster.";
L     Biol. Reprod. 53:454-461(1995).
N     Prorelaxin precursor.
N     RLN.
S     Mesocricetus auratus (Golden hamster).
C     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C     Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
C     Mesocricetus.
X     NCBI_TaxID=10036;
N     [1]
P     SEQUENCE FROM N.A.
C     TISSUE=Placenta;
X     MEDLINE=96115021; PubMed=7492700;
A     McCaslin R.B., Renegar R.H.;
T     "Determination of the prorelaxin nucleotide sequence and expression
T     of prorelaxin messenger ribonucleic acid in the golden hamster.";
L     Biol. Reprod. 53:454-461(1995).
-!-   FUNCTION: RELAXIN IS AN OVARIAN HORMONE THAT ACTS WITH ESTROGEN TO
      PRODUCE DILATATION OF THE BIRTH CANAL IN MANY MAMMALS. IT BEARS
      MATURE YOUNG, AND ALLOWS SEPARATION OF THE PELVIC BONES.
-!-   SUBUNIT: HETRODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
      DISULFIDE BONDS.
-!-   SUBCELLULAR LOCATION: Secreted.
-!-   SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
```

```
CC      PROMOTING EFFECTS OF THE IGFS ON CELL CULTURE. THEY ALTER THE
CC      INTERACTION OF IGFS WITH THEIR CELL SURFACE RECEPTORS.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- PTM: O-LINKED GLYCANS CONSIST OF HEX (PROBABLY GAL), HEXNAC
CC      (PROBABLY GALNAC) AND SIALIC ACID RESIDUES. MAJOR GLYCOFORMS
CC      CONSIST OF 8-16 MONOSACCHARIDES (BY SIMILARITY TO IGFBP-6
CC      EXPRESSED RECOMINANTLY IN CHO CELLS).
CC   -!- SIMILARITY: Contains 1 IGFBP domain.
CC   -!- SIMILARITY: Contains 1 thyroglobulin type-I domain.
CC      --------------------------------------------------------------
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      --------------------------------------------------------------
DR   EMBL; M62402; AAB06187.1; -.
DR   EMBL; AJ006952; CAA07346.1; -.
DR   EMBL; BC003507; AAH03507.1; -.
DR   EMBL; BC005007; AAH05007.1; -.
DR   EMBL; BC010162; AAH10162.1; -.
DR   EMBL; BC011708; AAH11708.1; -.
DR   EMBL; M69054; AAA88070.1; -.
DR   PIR; A39842; A39842.
DR   HSSP; P24593; 1BOE.
DR   Genew; HGNC:5475; IGFBP6.
DR   MIM; 146735; -.
DR   GO; GO:0008285; P:negative regulation of cell proliferation; TAS.
DR   InterPro; IPR000867; Insl_gro_fac_pr.
DR   InterPro; IPR000716; Thyroglobulin_1.
DR   Pfam; PF00219; IGFBP; 1.
DR   Pfam; PF00086; thyroglobulin_1; 1.
DR   SMART; SM00121; IB; 1.
DR   SMART; SM00211; TY; 1.
DR   PROSITE; PS00222; IGF_BINDING; FALSE_NEG.
DR   PROSITE; PS00484; THYROGLOBULIN_1; 1.
KW   Growth factor binding; Signal; Glycoprotein; Polymorphism.
FT   SIGNAL        1     24       PROBABLE.
FT   CHAIN        25    240       INSULIN-LIKE GROWTH FACTOR BINDING
FT                                PROTEIN 6.
FT   DOMAIN      186    234       THYROGLOBULIN TYPE I.
FT   DISULFID     29     44
FT   DISULFID     40     44
FT   DISULFID     57     63
FT   DISULFID     71     84
FT   DISULFID     78    104
FT   DISULFID    163    190
FT   DISULFID    201    212
FT   DISULFID    214    234
FT   CARBOHYD    126    126       O-LINKED (HEXNAC...).   (BY SIMILARITY).
FT   CARBOHYD    144    144       O-LINKED (HEXNAC...).   (BY SIMILARITY).
FT   CARBOHYD    145    145       O-LINKED (HEXNAC...).   (BY SIMILARITY).
FT   CARBOHYD    146    146       O-LINKED (HEXNAC...).   (BY SIMILARITY).
FT   CARBOHYD    152    152       O-LINKED (HEXNAC...).   (BY SIMILARITY).
FT   VARIANT     236    236       T -> P (IN dbSNP:1053134).
FT                                /FTId=VAR_011907.
FT   CONFLICT      2      2       T -> C (IN REF. 4).
SQ   SEQUENCE   240 AA;  25322 MW;  2853D231C025009 CRC64;

Alignment Scores:
Pred. No.:                                 1.78
Score:                                    80.00
Percent Similarity:                       44.95%      Length:         240
Best Local Similarity:                    28.44%      Matches:         31
Query Match:                              10.34%      Conservative:    18
DB:                                                   Mismatches:      40
                                                      Indels:          20
                                                      Gaps:             6

US-09-518-842-1 (1-420) x IB6_HUMAN (1-240)

Qy    25  CTGCCAGCAATCTGGCTGCTGAGCCAACTCCTTAGAGAAAGCCTAGCAGCAGAGCTG  84
          ||||||   |||   ||| |||||||||:|   ||||||||||::::   |||   |||
```

N   (2)
P   SEQUENCE FROM N.A.
X   MEDLINE=99189142; PubMed=10087296;
A   Ehrenborg E., Zazzi H., Lagercrantz S., Granqvist M., Hillerbrand U.,
A   Allander S.V., Larsson C., Luthman H.;
T   "Characterization and chromosomal localization of the human insulin-
T   like growth factor-binding protein 6 gene.";
L   Mamm. Genome 10:376-380(1999).
N   (3)
P   SEQUENCE FROM N.A.
C   TISSUE=Pancreas;
X   MEDLINE=22388257; PubMed=12477932;
A   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
A   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
A   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
A   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
A   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
A   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
A   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
A   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
A   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
A   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
A   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
A   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
A   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
A   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
A   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
A   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
A   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
T   "Generation and initial analysis of more than 15,000 full-length
T   human and mouse cDNA sequences.";
L   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
N   (4)
P   SEQUENCE OF 2-240 FROM N.A.
C   TISSUE=Placenta;
X   MEDLINE=91207395; PubMed=1719383;
A   Shimasaki S., Gao L., Shimonaka M., Ling N.;
T   "Isolation and molecular cloning of insulin-like growth factor-binding
T   protein-6.";
L   Mol. Endocrinol. 5:938-948(1991).
N   (5)
P   PARTIAL SEQUENCE OF 28-42.
X   MEDLINE=91207395; PubMed=1850257;
A   Andress D.L., Birnbaum R.S.;
T   "A novel human insulin-like growth factor binding protein secreted by
T   osteoblast-like cells.";
L   Biochem. Biophys. Res. Commun. 176:213-218(1991).
N   (6)
P   PRELIMINARY SEQUENCE OF 28-42.
C   TISSUE=Cerebrospinal fluid;
X   MEDLINE=90005986; PubMed=2551732;
A   Roghani M., Hossenlopp P., Lepage P., Balland A., Binoux M.;
T   "Isolation from human cerebrospinal fluid of a new insulin-like
T   growth factor-binding protein with a selective affinity for IGF-II.";
L   FEBS Lett. 255:253-258(1989).
N   (7)
P   CARBOHYDRATE-LINKAGE SITES.
X   MEDLINE=98241382; PubMed=9572875;
A   Neumann G.M., Marinaro J.A., Bach L.A.;
T   "Identification of O-glycosylation sites and partial characterization
T   of carbohydrate structure and disulfide linkages of human
T   insulin-like growth factor binding protein 6.";
L   Biochemistry 37:6572-6585(1998).
N   (8)
P   DISULFIDE BONDS.
X   MEDLINE=9926 2603; PubMed=10329650;
A   Neumann G.M., Bach L.A.;
T   "The N-terminal disulfide linkages of human insulin-like growth
T   factor-binding protein-6 (hIGFBP-6) and hIGFBP-1 are different as
T   determined by mass spectrometry.";
L   J. Biol. Chem. 274:14587-14594(1999).
CC   -!- FUNCTION: IGF-BINDING PROTEINS PROLONG THE HALF-LIFE OF THE IGFS
CC   AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH

```
Db    7  LeuProProLeuLeuLeuLeuLeuLeuAlaLeuLeuLeuLeuAlaAlaSerProGlyGlyAlaLeu  26

Qy   85  AGGGGATGTGGTCCCGATTTGGAAAAACACTTGCTGTCATATTGCCCATGCCTGAGAAG  144
                        ||||| |||  ||||::: :::  |||
Db   27  AlaArgCys---ProGlyCysGlyGlnGlyValGlnAlaGlyCys--------------  40

Qy  145  ACATTCACCACCACCCCCAGGAGGGGGGCCTGGCTGCTGAA------TCTGGACGTCCCAAAGAA  198
                   ||||||||| :::  |||       |||
Db   41  ---------------ProGlyGlyCysValGluGluGluAspGlyGlySerProAlaGlu  55

Qy  199  ATGGTGTCAACCTCCAAC-------AACAAAGATGGACAAGCCTTAGGTACGACATCA  249
                          |||     :::::||||||  |||
Db   56  GlyCysAlaGluAlaGlyGluCysLeuArgArgGluArgGlyGlnGluCysGlyVal------  73

Qy  250  GAATTCATTCCTAATTTGTCACCAGAGCTGAAG---AAACCACTGTCTGAAGGGCAGCCA  306
                 :::  ||||||| :::||||       |||  :::
Db   74  ---TyrThrProAsnCysAlaProGlyLeuGlnCysHisProProLysHisProLysAspGluAla  92

Qy  307  TCATTGAAGAAAATATATACTTTCCCGC  333
                |||::: ::::::||||  |||
Db   93  ProLeuArgAlaLeuLeuGlyArg  101

RESULT 15
NUOF_BUCAP
D   NUOF_BUCAP  STANDARD;  PRT;  447 AA.
C   Q8K9Y3;
DT  28-FEB-2003 (Rel. 41, Created)
DT  28-FEB-2003 (Rel. 41, Last sequence update)
DT  28-FEB-2003 (Rel. 41, Last annotation update)
DE  NADH-quinone oxidoreductase chain F (EC 1.6.99.5)  (NADH dehydrogenase
DE  I, chain F)  (NDH-1, chain F).
GN  NUOF OR BUSG151.
OS  Buchnera aphidicola (subsp. Schizaphis graminum).
OC  Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC  Enterobacteriaceae; Buchnera.
OX  NCBI_TaxID=98794;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=22084549; PubMed=12089438;
RA  Tamas I., Klasson L., Canbaeck B., Naeslund A.K., Eriksson A.-S.,
RA  Wernegreen J.J., Sandstroem J.P., Moran N.A., Andersson S.G.E.;
RT  "50 million years of genomic stasis in endosymbiotic bacteria.";
RL  Science 296:2376-2379(2002).
CC  -!- FUNCTION: NDH-1 shuttles electrons from NADH, via FMN and iron-
CC      sulfur (Fe-S) centers, to quinones in the respiratory chain.
CC      Couples the redox reaction to proton translocation (for every two
CC      electrons transferred, four hydrogen ions are translocated across
CC      the cytoplasmic membrane), and thus conserves the redox energy in
CC      a proton gradient (By similarity).
CC  -!- CATALYTIC ACTIVITY: NADH + quinone = NAD(+) + quinol.
CC  -!- COFACTOR: FMN and one 4Fe-4S cluster (Probable).
CC  -!- SUBUNIT: Composed of 13 different subunits. Subunits nuoCD, E, F,
CC      and G constitute the peripheral sector of the complex (By
CC      similarity).
CC  -!- SIMILARITY: BELONGS TO THE COMPLEX I 51 kDA SUBUNIT FAMILY.
CC  --------------------------------------------------------------------------
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  --------------------------------------------------------------------------
R   EMBL; AE014090; AAM67719.1; -.
R   InterPro; IPR001949; Complex1_51K.
R   Pfam; PF01512; Complex1_51K; 1.
R   PROSITE; PS00644; COMPLEX1_51K_1; 1.
R   PROSITE; PS00645; COMPLEX1_51K_2; 1.
W   Oxidoreductase; NAD; Quinone; Flavoprotein; FMN; Iron-sulfur;
W   4Fe-4S; Complete proteome.
T   NP_BIND    61    70  NAD (BY SIMILARITY).
T   NP_BIND   174   221  FMN (BY SIMILARITY).
```

```
FT  METAL  352  352  IRON-SULFUR (4FE-4S) (POTENTIAL).
FT  METAL  355  355  IRON-SULFUR (4FE-4S) (POTENTIAL).
FT  METAL  358  358  IRON-SULFUR (4FE-4S) (POTENTIAL).
FT  METAL  399  399  IRON-SULFUR (4FE-4S) (POTENTIAL).
SQ  SEQUENCE  447 AA;  49864 MW;  D11DE0940D67AB51 CRC64;

Alignment Scores:
Pred. No.:                         1.98        Length:         447
Score:                            80.00        Matches:         41
Percent Similarity:               36.55%       Conservative:    12
Best Local Similarity:            28.28%       Mismatches:      56
Query Match:                      10.34%       Indels:          36
DB:                                   1        Gaps:             6

US-09-518-842-1 (1-420) x NUOF_BUCAP (1-447)

Qy   28  CCAGCAATCTGG---CTGCTGCTGAGCCAACTCCTTAGAGAAGCCTAGCAGCAGAGCTG  84
                              ||||||   ||
Db  260  ProGlyValTrpGluLeuProPheGlyThrThrAlaArgGluIleLeuGluAspTyrAla  279

Qy   85  AGGGGATGTGGTCCCCGATTTGGAAAACACTTGCTGTCATATTGCCCATGCCTGAGAAG  144
         ||||||                                         |||::: :::
Db  280  ArgGlyMet-----LysSerGlyLeuPheLeuLysSerTrp-----------------  291

Qy  145  ACATTCACCACCACCCCCAGGAGGG----------TGGCTGCTGGAATCTGGACGTCCCC  192
                                  ||||||||       ::: |||::: ||
Db  292  ---------------GlnProGlyGlyAlaGlyThrAspPheGluIleGluLysHisLeuAsp  307

Qy  193  AAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACA------  246
                 ||||            |||            |||||| |||:::
Db  308  LeuProMetAspPheThrSerIleAlaLysAlaAlaGlySerArgLeuGlyThrAlaIleAla  327

Qy  247  ----------------------------------------------TCAGAATTCATT  258
                                                          :::|||
Db  328  MetAlaValAspAsnLysThrAsnMetIleSerLeuValCysAsnIleGluLysPhePhe  347

Qy  259  CCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCATTGAAGAAA  318
                ||||           |||                            :::
Db  348  SerArgGluSerCysGlyLeuCysThrProCysArgGluGlyLeuProTrpIleValLys  367

Qy  319  ATAATACTTTCCCGCAAAAAGAGAAGTGGACGTCACAGATTTGATCCATTCTGTTGTGAA  378
                |||::: ::::::|||   :::|||:::  |||||
Db  368  IleLeuGluSerLeuGluLysLysGluGly---HisLysAsnAspValLysAsnLeuGlu  386

Qy  379  GTAATTGTGACGAT  393
                :::|||
Db  387  ArgLeuCysLeuAsp  391

Search completed: October  9, 2003, 12:15:52
Job time : 20.5 secs
```

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2003 Compugen Ltd.
```

M nucleic - protein search, using frame_plus_n2p model

un on: October 9, 2003, 12:12:26 ; Search time 31 Seconds
                              (without alignments)
                              2605.861 Million cell updates/sec

itle: US-09-518-842-1
erfect score: 774
equence: 1 ATGGCCAGCCGTGTTCCGGTC..........CAGTAAATTATGTACACATAG 420

coring table: BLOSUM62
           Xgapop 10.0 , Xgapext 0.5
           Ygapop 10.0 , Xgapext 0.5
           Fgapop 6.0 , Fgapext 7.0
           Delop   6.0 , Delext   7.0

earched: 283308 seqs, 96168682 residues

otal number of hits satisfying chosen parameters: 566616

inimum DB seq length: 0
aximum DB seq length: 2000000000
ost-processing: Minimum Match 0%
           Maximum Match 100%
           Listing first 45 summaries

atabase : PIR_76.*
           1: _pirl:*
           2: pir2:*
           3: pir3:*
           4: pir4:*

```
ommand line parameters:
MODEL=frame+_n2p.model -DEV=x1p
Q=/cgn2_1/USPTO_spool/US09518842/runat_09102003_111037_13078/app_query.fasta_1.583
DB=PIR_76 -QFMT=fastan -SUFFIX=pir -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
UNITS=Bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
USER=US09518842 @CGN_1_1_62 @runat_09102003_111037_13078 -NCPU=6 -ICPU=3
NO_MMAP -LARGEQUERY -NEG_SCORES=100 -WAIT -DSPBLOCK=100 -LONGLOG
DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

	esult No.	Score	Query Match	Length	DB	ID	Description
	1	180	23.3	182	1	RXPG	relaxin precursor
	2	151	19.5	185	1	S48082	relaxin precursor
	3	123.5	16.0	185	2	A60982	relaxin 2 precursor
	4	122.5	15.8	185	2	A34936	relaxin precursor
	5	120	15.5	186	1	RXRT	relaxin precursor
	6	119.5	15.4	185	1	A44559	relaxin precursor
	7	113.5	14.7	166	2	S42786	relaxin 2 precursor
	8	111.5	14.4	178	2	A49014	20K protein - rabb
	9	108.5	14.0	166	2	S42783	relaxin 1 precursor
	10	103.5	13.4	143	2	I47053	relaxin B, C and A
	11	101	13.0	560	2	T16833	hypothetical prote
	12	94	12.6	754	2	AE0614	probable competenc
c	13	83	10.7	321	2	E83259	hypothetical prote
	14	82.5	10.7	1626	2	T09271	probable tail comp
c	15	82	11.0	335	1	T02809	probable peptidylp
	16	80.5	10.4	457	2	AG0433	proteinase (EC 3.4
	17	80	10.3	240	2	A39842	insulin-like growt
c	18	77.5	10.4	1217	2	T25894	hypothetical prote
	19	77	9.9	1490	2	S72351	nonstructural poly
	20	76.5	9.9	160	2	A49194	relaxin - guinea p
c	21	76.5	10.3	402	2	A84581	probable disease r
	22	76	9.8	188	2	JC6547	high sulfur protei
	23	75.5	9.7	391	2	S60672	replication protei
	24	75	9.7	224	2	B40850	secretory protein
	25	75	10.1	993	2	S46779	26S proteasome reg
c	26	75	10.1	1711	1	A55148	protein-tyrosine-p
c	27	74.5	10.0	564	2	T38291	GATA-type transcri
c	28	74.5	10.0	564	2	T43298	transcription fact
c	29	74.5	10.0	1531	2	H71468	probable outer mem
	30	73.5	9.6	273	2	G89979	NAD synthetase, pr
c	31	73.5	9.9	296	1	OPNB7	peroxidase (EC 1.1
	32	73.5	9.5	453	2	I50674	retinoic acid rece
c	33	73	9.8	315	2	T10252	peroxidase (EC 1.1
	34	73	9.4	1435	2	S69632	regulatory protein
c	35	73	9.8	1744	1	C4HU	complement C4A pre
	36	73	9.4	2493	2	S26372	nonstructural poly
	37	73	9.4	2493	2	S72349	nonstructural poly
	38	72.5	9.4	292	2	AB0825	probable membrane
	39	72.5	9.4	1259	2	T16038	hypothetical prote
c	40	72	9.7	403	2	T20070	hypothetical prote
c	41	72	9.7	443	2	AI3525	4-aminobutyrate tr
	42	72	9.3	479	2	B87092	GTP1/Obg-family GT
	43	72	9.3	483	2	AD3618	succinate-semialde
	44	72	9.3	752	2	T50450	hypothetical prote
	45	72	9.3	836	1	JDVLD	DNA-directed DNA p

ALIGNMENTS

RESULT 1
RXPG
relaxin precursor - pig
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 02-Apr-1982 #sequence_revision 02-Apr-1982 #text_change 18-Jun-1999
C;Accession: A90934; A93187; A90205; A90196; S32312; A29796; A01615
R;Haley, J.; Hudson, P.; Scanlon, D.; John, M.; Cronk, M.; Shine, J.; Tregear, G.; Nial
DNA 1, 155-162, 1982
A;Title: Porcine relaxin: molecular cloning and cDNA structure.
A;Reference number: A90934; MUID:83157118; PMID:6897721
A;Accession: A90934
A;Molecule type: DNA
A;Residues: 1-182 <HAL>
A;Cross-references: GB;K01088; NID:g164634; PIDN:AAA31114.1; PID:g164635
R;James, R.; Niall, H.; Kwok, S.; Bryant-Greenwood, G.
Nature 267, 544-546, 1977
A;Title: Primary structure of porcine relaxin: homology with insulin and related growth
A;Reference number: A93187; MUID:77213067; PMID:876374
A;Accession: A93187
A;Molecule type: protein
A;Residues: 25-50,'TWGR',161-182 <JAM>
R;Schwabe, C.; McDonald, J.K.; Steinetz, B.G.
Biochem. Biophys. Res. Commun. 75, 503-510, 1977
A;Title: Primary structure of the B-chain of porcine relaxin.
A;Reference number: A90205; MUID:77157271; PMID:851452
A;Accession: A90205
A;Molecule type: protein
A;Residues: 25-47,'VW',50 <SCH1>
R;Schwabe, C.; McDonald, J.K.
Biochem. Biophys. Res. Commun. 74, 1501-1504, 1977
A;Title: Demonstration of a pyroglutamyl residue at the N terminus of the B-chain of po
A;Reference number: A90201; MUID:77134136; PMID:843375
A;Contents: annotation; pyrrolidone carboxylic acid
R;Schwabe, C.; McDonald, J.K.; Steinetz, B.G.
Biochem. Biophys. Res. Commun. 70, 397-405, 1976
A;Title: Primary structure of the A chain of porcine relaxin.
A;Reference number: A90196; MUID:76231539; PMID:938497
A;Accession: A90196

A;Molecule type: protein
A;Residues: 161-169,'E',171-182 <SCH>
R;Schwabe, C.; McDonald, J.K.
Science 197, 914-915, 1977
A;Title: Relaxin: a disulfide homolog of insulin.
A;Reference number: A94245; MUID:77236040; PMID:887933
A;Contents: annotation; disulfide bonds
R;Kohsaka, T.; Takahara, H.; Sugawara, K.; Tagami, S.
Biol. Chem. Hoppe-Seyler 374, 203-210, 1993
A;Title: Endogenous heterogeneity of relaxin and sequence of the major form in pregnant
A;Reference number: S32313; MUID:93257096; PMID:8489740
A;Accession: S32313
A;Molecule type: protein
A;Residues: 25-53 <KOH>
A;Accession: S32312
A;Molecule type: protein
A;Residues: 161-182 <KO2>
R;Haley, J.; Crawford, R.; Hudson, P.; Scanlon, D.; Tregear, G.; Shine, J.; Niall, H.
J. Biol. Chem. 262, 11940-11946, 1987
A;Title: Porcine relaxin. Gene structure and expression.
A;Reference number: A29796; MUID:87308187; PMID:2442155
A;Accession: A29796
A;Molecule type: DNA
A;Residues: 1-115, 'L',117-182 <HA2>
A;Cross-references: GB:J02792; NID:g164636; PIDN:AAA31115.1; PID:g164637
C;Comment: Relaxin is an ovarian hormone that acts with estrogen to produce dilation of
C;Genetics:
A;Introns: 70/1
C;Superfamily: insulin
C;Keywords: Pyroglutamic acid
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-56/Domain: relaxin chain B #status experimental <RCB>
F;25-56,161-182/Product: relaxin #status experimental <MAT>
F;161-182/Domain: relaxin chain A #status experimental <RCA>
F;161-182/Domain: relaxin chain A (Gln) (in mature form) #status experimen
F;25/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experimental
F;34-169,46-182,168-173/Disulfide bonds: #status experimental

Alignment Scores:
Pred. No.:
Score: 6.44e-11   Length: 182
                          180.00   Matches: 61
Percent Similarity: 38.02%   Conservative: 12
Best Local Similarity: 31.77%   Mismatches: 55
Query Match: 23.26%   Indels: 64
DB: 1   Gaps: 6

US-09-518-842-1 (1-420) x RXPG (1-182)

Qy   1 ATGGCCAGCCTGTTCCGGTCCTATCTGCCAGCAATCTGGCTGCTGCTGAGCCAACTCCTT 60
            || ||||| |||||| ||||||||||||||||
Db   1 MetProArgLeuPhe---SerTyrLeuLeuGlyValTrpLeuLeuSerGlnLeuPro 19

Qy   61 AGAGAA------------AGCCTAGCAGCAGCAGAGCTGAGGGGATGGTGTCCCGATTTGGA 108
            |||||| || |||||| :::: ||||
Db   20 ArgGluIleProGlyGlnSerThrAsnAspPheIleLysAlaCysPheLysGluGluLeuVal 39

Qy   109 AAACACTTGCTGTGTCATATTGCCCCATGCCTGAGAAGACATTCACCACCACCAGGAGGG 168
            ::: ||||||| ||::: |:::
Db   40 ArgLeuTrpValGlyIleCysGlySerVal---------------------Ser 50

Qy   169 TGG-----------------------CTGCTGGAATCTGGACGTCCCAAAGAA 198
            ||| |||::|| ||||||
Db   51 TrpGlyArgThrAlaLeuSerLeuGluGluGlnLeuProGlnLeuProAlaGlu 70

Qy   199 ATGGTGTCAACCTCCAACAACAACAAGATGACAAGCCTTAGGTACGACCATCAGAATTCATT 258
            :::   ::||||| ||||| ||| |||:::|:::
Db   71 ThrMetProSerIleThrLysAspAlaGlnIleLeuLysMetMetLeuGluPheVal 90

Qy   259 CCTAATTTGTCACCAGAGCTGAAGAACCACTGTCTGAAGGGCAGCCATCATTG------ 312
            |||||||||||| |||||||||||||||||||||||||||
Db   91 ProAsnLeuProGlnLeuGluLeuLysAlaThrLeuSerGluArgGlnProSerLeuArgGlu 110

Qy   313 -----------------------AAGAAAATAATA 324
                                   ||||||||||||

Db   111 LeuGlnGlnSerAlaSerLysAspSerAsnLeuAsnPheGluPheLysLysIleIle 130
Qy   325 CTT------------------------ 327
            |||
Db   131 LeuAsnArgGlnAsnGluAlaGluAspLysSerLeuLeuGluLeuLysAsnLeuGlyLeu 150
Qy   328 ------------TCCCGCAAAAGAAGAAGTGGACGTCACAGATTTGATCCATTCTGTTGTGAA 378
                        |||||||||||||||||| ||||||||| |:::
Db   151 AspLysHisSerArgLysLysArgLeuPheArgMetThrLeuArgGluCysLysCysCysGln 170
Qy   379 GTAATTTGTGACGATGGAACTTCAGTAAATTATGT 414
            ||| ||| :::|||||||
Db   171 ValGlyCysIleArgLysAspIleAlaArgLeuCys 182

RESULT 2
S48082
relaxin precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 07-May-1995 #sequence revision 21-Jul-1995 #text_change 16-Jul-1999
A;Accession: S48082; PC2067; PN0626
R;Evans, B.A.; John, M.; Fowler, K.J.; Summers, R.J.; Cronk, M.; Shine, J.; Tregear, G.
J. Mol. Endocrinol. 10, 15-23, 1993
A;Title: The mouse relaxin gene: nucleotide sequence and expression.
A;Reference number: S48082; MUID:93199663; PMID:8452637
A;Accession: S48082
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-185 <EVA>
A;Cross-references: EMBL:Z27088; NID:g414780; PIDN:CAA81611.1; PID:g414781
R;Buellesbach, E.E.; Schwabe, C.
Biochem. Biophys. Res. Commun. 196, 311-319, 1993
A;Title: Mouse relaxin: synthesis and biological activity of the first relaxin with an
A;Reference number: PN0626; MUID:94030011; PMID:8216305
A;Accession: PC2067
A;Molecule type: protein
A;Residues: 23-57 <BUE>
A;Accession: PN0626
A;Molecule type: protein
A;Residues: 161-185 <BU2>
A;Note: proteins with and without 184-Tyr were synthesized, their biological activities
C;Superfamily: insulin
C;Keywords: hormone
F;23-57,161-185/Product: relaxin #status experimental <MAT>
F;23-57/Domain: chain B #status experimental <CHB>
F;161-185/Domain: chain A #status experimental <CHA>
F;36-171,48-185,170-175/Disulfide bonds: #status experimental

Alignment Scores:
Pred. No.:   Length: 185
Score: 8.34e-08   Matches: 50
                          151.00   Conservative: 19
Percent Similarity: 37.70%   Mismatches: 68
Best Local Similarity: 27.32%   Indels: 46
Query Match: 19.51%   Gaps: 3
DB: 2

US-09-518-842-1 (1-420) x S48082 (1-185)

Qy   1 ATGGCCAGCCTGTTCCGGTCCTATCTGCCAGCAATCTGGCTGCTGCTGAGCCAACTCCTT 60
            |||::|||| ||| ||||||||||||||||||||||||
Db   1 MetSerSerArgPheLeuLeuGlnLeuLeuGlyPheTrpLeuLeuSerGlnLeuProCys 20

Qy   61 AGAGAAAGCCTAGCAGCAGAG-----------------CTGAGGGGATGGTGTCCCGATTT 105
            ||| ::: ||||||
Db   21 ArgThrArgValSerGluGluTrpMetAspGlyPheIleAlaArgMetCysGlyArgGluTyr 40

Qy   106 GGAAAACACTTGCTGTGTCATATTGCCCCATGCCTGAGAAGACATTCACCACCACCAGGA 165
            ::: |||||||
Db   41 AlaArgGluLeuLeuIleLysCysGlyValGlyAlaSerValGlyArgLeuAlaLeuSerGlnGlu 60

Qy   166 GGGTGGCTGCTGGAATCTGGACGTCCCAAAGAAATGGTGTCAACCTCCAACAACAACAAGAT 225
            ||| ::: |:::   :::
Db   61 GluProAlaLeuLeuLeuAlaArgGlnAlaThrGluAsnPheIleAsnLysAsnLysAsp 80

Left column:

```
Qy    226 GGACAAGCCTTAGGTACGACATCAGAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAA 285
              |||||       :::  :: |||||||||||| |||||||||
Db     81 AlaGluProPheAspThrThrLeuLysCysLeuProAsnLeuSerGluGluLeuLysAla 100

Qy    286 CCACTGCTGAAGGGCAGCCATCATTGAAGAAAATA--------------ATACTT------ 327
              |||||                                    ::|||
Db    101 ValLeuSerGluAlaGlnAlaSerLeuLeuProGluLeuGlnHisAlaProValLeuSerSerAsp 120

Qy    327
Db    327

Db    121 SerValValSerLeuGluGlyPheLysLysThrLeuHisAspArgLeuGlyGluAlaGlu 140

Qy    328 ------------------------------------------TCCCGCAAAAAGAGA 342
                                                       |||||||||||
Db    141 AspGlySerProProGlyLeuLysTyrLeuGlnSerAspThrHisSerArgLysLysArg 160

Qy    343 AGTGGACGTCACAGATTTGATCCATTCGTGTTGTGAAGTAATTTGTGACGATGGAACTTCA 402
              ||||||||| |||  ||||||
Db    161 GluSerGlyGlyLeuMetSerGlnGlnCysCysHisValGlyCysSerArgSerIle 180

Qy    403 GTTAAAATTA 411
              ||||||||
Db    181 AlaLysLeu 183
```

```
RESULT 3
A60982
relaxin 2 precursor [validated] - human
;Alternate names: preprorelaxin 2
;Species: Homo sapiens (man)
;Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 08-Dec-2000
;Accession: A05092, A60982
;Hudson, P.; John, M.; Crawford, R.; Haralambidis, J.; Scanlon, D.; Gorman, J.; Tregear
  EMBO J. 3, 2333-2339, 1984
;Title: Relaxin gene expression in human ovaries and the predicted structure of a human
;Reference number: A05092; MUID:85051298; PMID:6548702
;Accession: A05092
;Molecule type: mRNA
;Residues: 1-185 <HUD>
;Cross-references: GB:X00948; NID:g35926; PIDN:CAA25460.1; PID:g35927
;Stults, J.T.; Bourell, J.H.; Canova-Davis, E.; Ling, V.T.; Laramee, G.R.; Winslow, J.W
  Biomed. Environ. Mass Spectrom. 19, 655-664, 1990
;Title: Structural characterization by mass spectrometry of native and recombinant huma
;Reference number: A60982; MUID:91167739; PMID:2076464
;Accession: A60982
;Molecule type: protein
;Residues: 25-53;162-185 <STU>
;Genetics:
;Gene: GDB:RLN2
;Cross-references: GDB:119553; OMIM:179740
;Map position: 9pter-9q12
;Superfamily: insulin
;Keywords: ovary; pyroglutamic acid
;1-24/Domain: signal sequence #status predicted <SIG>
;25-53/Domain: relaxin 2 chain B #status experimental <BCH>
;25-53;162-185/Product: relaxin 2 #status experimental <MAT>
;58-157/Domain: relaxin 2 connecting C peptide #status predicted <CPEP>
;162-185/Domain: relaxin 2 chain A #status experimental <ACH>
;35-172,47-185,171-176/Disulfide bonds: #status experimental
;162/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experime
```

```
Alignment Scores:
Pred. No.:                               7.45e-05     Length:         185
Score:                                   123.50       Matches:        53
Percent Similarity:                      35.35%       Conservative:   17
Best Local Similarity:                   26.77%       Mismatches:     55
Query Match:                             15.96%       Indels:         73
DB:                                                   Gaps:           7

US-09-518-842-1 (1-420) x A60982 (1-185)

Qy      1 ATGGCCAGCCGTGTTCCGGTCCTATCTGCCAGCAATCTGCCTGCTGCTGAGCCAACTCCTT 60
            ||  ||                          :::  ||||||||:::|||
Db      1 MetProArgLeuPhePhePheHisLeuLeuLeuGlyValCysLeuLeuLeuAsnGlnPheSer 20
```

Right column:

```
Qy     61 ---------------------------------------------AGAGAAAGCCTA 72
                                                          |||||| :::
Db     21 ArgAlaValAlaAspSerTrpMetGluGluValIleLysLeuCysGlyArgGluLeuVal 40

Qy     73 GCAGCAGCTGAGGGGATGGGT---CCCCGATTGGAAAACACTTGCTGTCATATTGC 129
            |||::::::                         :::   |||  |||||||||
Db     41 ArgAlaGlnIleAlaIleCysGlyMetSerThrTrpSerLysSerLeuSerGlnGlu 60

Qy    130 CCCATGCTGAGAAGACATTCACCACCACCACCCAGGAGGGTGCGTCGTGGAATCTGGACGT 189
            ||||||                   ::::              |||
Db     61 AspAlaProGln----------------ThrPro--------------Arg 67

Qy    190 CCC---AAAGAAATGTGTCAACCTCCAACAACAAAGATGACAAGCCTTAGGTACGACA 246
            |||                |||||||||||       :::
Db     68 ProValAlaGluLeuValProSerPheIleAsnLysAspThrGluThrIleAsnMetMet 87

Qy    247 TCAGAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCA 306
            ||||||||| |||  ||||||
Db     88 SerGluPheValAlaAsnLeuProGlnGluLeuLysLeuThrLeuSerGluMetGlnPro 107

Qy    307 TCATTG- 312
            ::::|||
Db    108 AlaLeuProGlnLeuGlnHisValProValLeuLysAspSerSerLeuLeuPheGlu 127

Qy    313 ------AAGAAAATAATACTT--------- 327
            ::::|||
Db    128 GluPheLysLysLeuIleArgAsnArgGlnSerGluAlaAlaAspSerSerProSerGlu 147

Qy    328                      TCCCGCAAAAAGAGAAGAAGLysArgGlnLeuTyrSerAlaLeu 360
                                 |||||||||||||
Db    148 LeuLysTyrLeuGlyLeuAspThrHisSerArgLysArgGlnLeuTyrSerAlaLeu 167

Qy    361 GATCCATTCGTGTTGTGAAGTAATTGTGACGATGGAACTTCAGTTAAATTATGT 414
            |||               :::   |||
Db    168 AlaAsnLysCysCysHisValGlyCysThrArgSerLeuAlaArgPheCys 185
```

```
RESULT 4
A34936
relaxin precursor - rhesus macaque
;Species: Macaca mulatta (rhesus macaque)
;Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 12-Jun-1998
;Accession: A34936
;Crawford, R.J.; Hammond, V.E.; Roche, P.J.; Johnston, P.D.; Tregear, G.W.
  J. Mol. Endocrinol. 3, 169-174, 1989
;Title: Structure of rhesus monkey relaxin predicted by analysis of the single-copy rh
;Reference number: A34936; MUID:90073957; PMID:2590381
;Accession: A34936
;Status: not compared with conceptual translation
;Molecule type: mRNA
;Residues: 1-185 <CRA>
;Superfamily: insulin
;1-22/Domain: signal sequence #status predicted <SIG>
;23-185/Product: relaxin #status predicted <MAT>
```

```
Alignment Scores:
Pred. No.:                               9.53e-05     Length:         185
Score:                                   122.50       Matches:        51
Percent Similarity:                      34.34%       Conservative:   17
Best Local Similarity:                   25.76%       Mismatches:     57
Query Match:                             15.83%       Indels:         73
DB:                                                   Gaps:           6

US-09-518-842-1 (1-420) x A34936 (1-185)

Qy      1 ATGGCCAGCCGTGTTCCGGTCCTATCTGCCAGCAATCTGCCTGCTGCTGAGCCAACTCCTT 60
            ||  ||                          :::  ||||||||:::|||
Db      1 MetProArgLeuPheLeuPheHisLeuLeuLeuGlyValCysLeuLeuLeuAsnGlnPheSer 20

Qy     61 ---------------------------------------------AGAGAAAGCCTA 72
                                                          |||||| :::
Db     21 ArgAlaValAlaAlaAlaAlaTrpMetAspAspValIleLysAlaCysGlyArgGluLeuVal 40
```

```
Qy  73 GCAGCGAGCTGAGGGATGTGGT--CCCCGATTTGGAAAACACTTGCTGTCATAATTGC 129
             ||| :::::::  |||||
Db  41 ArgAlaGlnIleAlaIleCysGlyLysGlyLysSerThrLeuGlyLysLysArgSerLeuAsnGln:Glu 60

Qy 130 CCCATGCTGAGAAGACATTCACCACCACCCCCAGGAGGGTGGCTGCTGGAATCTGGACGT 189
             ||||||||::: :::                                   |||
Db  61 AspAlaProLeuLysPro------------------------------------Arg 67

Qy 190 CCC--AAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACA 246
             |||  :::|||
Db  68 ProAlaAlaGluIleValProSerLeuIleAsnAspThrGluThrIleAsnMetMet 87

Qy 247 TCAGAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCCAGCCA 306
             ||||||||:::  ||  |||
Db  88 SerGluPheValAlaAlaAsnLeuProGlnGluLeuLysLeuThrLeuSerGluGluArgGlnPro 107

Qy 307 TCATTG---------------------------------------- 312
             ::|||
Db 108 AlaLeuSerGluGluLeuGlnGlnHisValProValLeuLeuLysAspSerAsnLeuSerPheGlu 127

Qy 313 ------AAGAAAATATACTT------------------------ 327
             ||||||:::
Db 128 GluPheLysLysIleIleIleArgArgGlnSerGluAlaThrAspSerSerProSerGlu 147

Qy 328 ------------------------TCCCGCAAAAAGAAGAGGACGTCACAGATT 360
                                    ::||||||:::|||||
Db 148 LeuArgSerLeuGlyLeuAspThrHisSerArgArgLysGlnLeuTyrMetThrLeu 167

Qy 361 GATCCATTCTGTTGTGAAGTAATTGTGACGATGGAACTTCAGTTAAATTATGT 414
             ||||  |||  :::  |||
Db 168 SerArgAsnLysCysCysHisIleGlyCysThrLysThrLysLysSerLeuSerLeuAlaLysPheCys 185
```

RESULT 5
RXRT
relaxin precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 18-Dec-1981 #sequence_revision 18-Dec-1981 #text_change 24-Sep-1999
C;Accession: A01614
R;Hudson, P.; Haley, J.; Cronk, M.; Shine, J.; Niall, H.
Nature 291, 127-131, 1981
A;Title: Molecular cloning and characterization of cDNA sequences coding for rat relaxin
A;Reference number: A01614; MUID:81197624; PMID:7231533
A;Accession: A01614
A;Molecule type: mRNA
A;Residues: 1-186 <HUD>
A;Cross-references: GB:J00780; GB:M25468; NID:g206606; PIDN:AAA42029.1; PID:g206607
C;Superfamily: insulin
C;Keywords: hormone; ovary; pyroglutamic acid
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-57/Domain: relaxin chain B #status predicted <RXB>
F;23-57,163-186/Product: relaxin #status predicted <MAT>
F;58-162/Domain: relaxin connecting C peptide #status predicted <RXC>
F;163-186/Domain: relaxin chain A #status predicted <RXA>
F;36-173,48-186,172-177/Disulfide bonds: #status predicted
F;163/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicte

```
lignment Scores:
red. No.:
core:                    0.000177      Length:          186
ercent Similarity:       120.00        Matches:          46
est Local Similarity:     33.16%       Conservative:     18
uery Match:               23.83%       Mismatches:       67
B:                        15.50%       Indels:           62
                                       Gaps:              5

S-09-518-842-1 (1-420) x RXRT (1-186)

y   1 ATGGCCAGCCGTTCCGGTCCTATCTGCCAGCAATCTGGCTGCTGCTGAGCCAACTCCTT 60
           |||  :::|||
b   1 MetSerSerArgSerArgLeuLeuGluLeuLeuLeuGlyPheTrpLeuPheLeuSerGlnProCys 20

y  61 ACAGAAAGCCTAGCAGCAGAG------------CTGAGGGGATGTGTCCCGATTT 105
           |||  :::|||
b  21 ArgAlaArgValSerGluGluTrpMetAspGlnValIleGluInValCysGlyArgGlyTyr 40
```

```
Qy 106 GGAAAACACTTGCTGTCATATTGCCCATGCTGAGAGAGACATTCACCACCCAGGA 165
             :::  |||
Db  41 AlaArgAlaTrpIleGluValCys----------GlyAlaSerValGly 53

Qy 166 GGGTGGCTGCTGGAATCTGACGTCC------------------AAAGAAATG 201
             |||||||||||:::                      |||: :::
Db  54 ArgLeuAlaLeuSerGlnGluGluProAlaProLeuAlaArgGlnAlaThrGluVal 73

Qy 202 GTGTCAACCTCCAACACAACAAAGATGGACAAGCCTTAGGTACGACATCAGAATTCATTCCT 261
             |||  ||| ||||||||      ::  :: :|||
Db  74 ValProSerPheIleAsnLysAspAlaGluProPheAspMetThrLeuLysCysLeuPro 93

Qy 262 AATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCCAG---------- 303
             ||||||||| ||| ||||||||||||||:::
Db  94 AsnLeuSerGluGluArgLysGluAlaAlaLeuSerGluGluGlyPheGluProGluLeu 113

Qy 303 TCATTG---------------------------------------- 303
             ::|||
Db 114 GlnGlnHisAlaProAlaAlaLeuSerAspSerValValSerLeuGluGlyPheLysLysThr 133

Qy 304 ------------------------CCATCATTGAAGAAAATA--- 321
                                     ||||||||
Db 134 PheHisAsnGlnLeuGlyAlaGluAspGlyGlyLeuProProGluLeuLysTyrLeuGly 153

Qy 322 ------------ATACTTTCCCGCAAAAAGAAGAGAAGTGGACGTCACAGATTTGATCCATTCTGTTGT 375
                         ||||||||||||||
Db 154 SerAspAlaGlnSerArgLysLysGlnSerArgLysLysGlnAlaAlaLeuLeuSerGluGluCysCys 173

Qy 376 GAAGTAATTGTGACGATGGAACTTCAGTTAAATTATGT 414
             ||||||||
Db 174 HisIleGlyCysThrArgArgSerIleAlaLysLeuAlaLeuLeuCys 186
```

RESULT 6
A44559
relaxin 1 precursor - human
N;Alternate names: preprorelaxin 1
C;Species: Homo sapiens (man)
C;Date: 03-Feb-1994 #sequence_revision 07-Oct-1994 #text_change 18-Jun-1999
C;Accession: B05092; A44559
R;Hudson, P.; John, M.; Crawford, R.; Haralambidis, J.; Scanlon, D.; Gorman, J.; Tregea
A;Title: Relaxin gene expression in human ovaries and the predicted structure of a huma
EMBO J. 3, 2333-2339, 1984
A;Reference number: A05092; MUID:85051298; PMID:6548702
A;Accession: B05092
A;Molecule type: DNA
A;Residues: 1-185 <HU1>
A;Cross-references: GB:X00949; NID:g35912; PIDN:CAA25461.1; PID:g35933
R;Hudson, P.; Haley, J.; John, M.; Cronk, M.; Crawford, R.; Haralambidis, J.; Tregear,
Nature 301, 628-631, 1983
A;Title: Structure of a genomic clone encoding biologically active human relaxin.
A;Reference number: A44559; MUID:83141755; PMID:6298628
A;Accession: A44559
A;Molecule type: DNA
A;Residues: 1-185 <HU2>
A;Cross-references: GB:X00949; NID:g35912; PIDN:CAA25461.1; PID:g35933
C;Comment: Relaxin is an ovarian hormone that acts with estrogen to produce dilation of
    disulfide bonds.
C;Genetics:
A;Gene: GDB:RLN1
A;Cross-references: GDB:119552; OMIM:179730
A;Map position: 9pter-9q12
C;Superfamily: insulin
C;Keywords: hormone; ovary
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-57/Domain: relaxin 1 chain B #status predicted <BCH>
F;26-57,162-185/Product: relaxin 1 #status predicted <MAT>
F;58-158/Domain: relaxin 1 connecting C peptide #status predicted <CPEP>
F;162-185/Domain: relaxin 1 chain A #status predicted <ACH>
F;35-172,47-185,171-176/Disulfide bonds: #status predicted

```
Alignment Scores:
Pred. No.:                    0.0002      Length:          185
```

RESULT 7 (continued)

Score: 119.50
Percent Similarity: 37.11%
Best Local Similarity: 26.80%
Query Match: 15.44%

B: Matches: 52   Conservative: 20   Mismatches: 57   Indels: 65   Gaps: 7

US-09-518-842-1 (1-420) x A44559 (1-185)

```
Qy   1 ATGGCCAGCCTGTTCCGGTCCTATCTGCCAGCAATCTGCTCGTGCTGAGCCAAACTCCTT  60
         |||||  ||||||                          :::||
Db   1 MetProArgLeuPheLeuHisLeuLeuLeuGluPheCysLeuLeuAsnGlnPheSer      20

Qy  61 AGAGAAAGCCTAGCAGCAGAG----CTGAGGGGGATGTGGTGTCCCCGATTT         105
         ||| :::||:|||||||                            |||||
Db  21 Arg--AlaValAlaAlaLysTrpLysAspValIleLeuLysLeuAsnCysGlyArgGluLeu  39

Qy 106 GGAAAACACTTGCTGTCATATTGCCCCATG--------CCTGAG                 141
         ::: |||||  |||                   :::
Db  40 ValValArgAlaGlnIleAlaIleCysGlyMetSerThrTrpSerLysArgSerLeuSerGln  59

Qy 142 AAGACATTCACCACCCCAGGAGGGTGCGTGCTGGCTGGAATCGGACGTCCC---AAAGAA  198
         |||||                           |||||
Db  60 GluAspAlaProGlnThrPro-----------------ArgProValAlaGlu        71

Qy 199 ATGGTGTCAACCTCCAACAACAAAGATGACAAGCCTAGGTACGACATCAGAGAATTCATT  258
         ||||||||    :::  ||||||||  ::: |||||||||||
Db  72 IleValProSerPheIleAsnLysAspThrGluThrIleIleIleMetLeuGluProIle   91

Qy 259 CCTAATTGTCACCAGAGTCGAGAAAACCACTGTCTGAAGGGCAGCCATCATTG------  312
         ||||||||||||||||||||||||||||||
Db  92 AlaAsnLeuProProGluLeuLysAlaAlaLeuSerGluArgGlnProSerLeuProGlu  111

Qy 313 --------AAGAAA                                               318
                :::
Db 112 LeuGlnGlnTyrValProAlaLeuProGlnLeuLysAspSerAsnLeuSerPheGluGluProLysLys 131

Qy 319 ATAATACTTTCCGC--------                                       333
         ::||| :::|||
Db 132 LeuIleLeuArgAsnArgGlnSerGluAlaAlaAlaAspSerAsnProSerGluLeuLysTyrLeu 151

Qy 334 ----------AAAAAGAGAAGTGACGTGACAGATTGATCCATTCTGT              372
                   ||||||||
Db 152 GlyLeuAspThrHisSerGlnLysLysArgArgProTyrValAlaLeuPheGluLysCys  171

Qy 373 TGTGAAGTAATTGTGACGATGGAACTTCAGTTAAATTATGT                    414
         |||  :::  ||||
Db 172 CysLeuIleGlyCysThrLysArgSerLeuAlaLysTyrCys                   185
```

RESULT 7
S42786
relaxin 2 precursor - chimpanzee (fragment)
C;Species: Pan troglodytes (chimpanzee)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 16-Jul-1999
G;Evans, B.B.
Submitted to the EMBL Data Library, November 1993
C;Reference number: S42786
C;Accession: S42786
C;Molecule type: mRNA
C;Residues: 1-166 <EVA>
C;Cross-references: EMBL:Z27245; NID:g416109; PIDN:CAA81758.1; PID:g416110
C;Genetics:
C;Gene: rlx2
C;Superfamily: insulin
A;1-5/Domain: signal sequence (fragment) #status predicted <SIG>
A;6-166/Product: relaxin 2 #status predicted <MAT>

Alignment Scores:
Pred. No.:
Score: 0.000871   Length: 166
Percent Similarity: 113.50   Matches: 45
Best Local Similarity: 36.42%   Conservative: 14
Query Match: 27.78%   Mismatches: 46
                        14.66%   Indels: 57

DB:

US-09-518-842-1 (1-420) x S42786 (1-166)

```
Qy  61 AGAGAAAGCCTAGCAGCAGCAGAGAGCTGAGGGGGATGTGGT---CCCCGATTTGGAAAACACTTG 117
         |||||| ::: ||| ::::::                      |||
Db  18 ArgGluLeuValArgAlaAlaGlnIleAlaIleCysGlyLysSerThrTrpSerLysArgSer    37

Qy 118 CTGTCATATTGCCCCATGCTGAGAAGACATTCACCACCCCAGGAGGGTGCGTGCTG        177
         |||||| ::: |||                              |||
Db  38 LeuSerGlnGluAspAlaProGln-----------------ThrPro--               47

Qy 178 GAATCTGGACGTCCC--AAAGAAATGGTGTCAACCTCCAACAACAAAGATGACAAGCC      234
         |||||                     ||||||||||    :::  ||||||||||
Db  48 -------ArgProValAlaGluIleValProSerPheIleAsnLysAspThrGluThr       64

Qy 235 TTAGTACGACATCAGAATTCATTCCTAATTGTCACCAGAGTCGAGAAAACCACTGTCT      294
         ::: ||||||||||||         |||||||  |||||||
Db  65 IleAsnMetSerGluPheValAlaAsnLeuProGlnGluLysLeuThrLeuSer          84

Qy 295 GAAGGGCAGCCATCATTG                                            312
         ||| ||||||||:::||
Db  85 GluMetGlnProAlaLeuProGlnLeuGlnTyrValProLeuLysAspSerSer         104

Qy 313 ------AAGAAAATAATACTT--                                       327
              ||:::::
Db 105 LeuLeuPheGluGluProLysLysLeuIleLeuArgAsnArgGlnSerGluAlaAlaAlaAspSer 124

Qy 328 --------TCCCGCAAAAGAGAAGTGA                                   348
                 |||||||||||||
Db 125 SerProSerGluLeuLysTyrLeuGlyLeuAspThrHisSerArgLysArgArgGlnLeu   144

Qy 349 CGTCACAGATTGATCCATTCTGTCGTGAAGTAATTGTGACGATGAGACTTCAGTTAAA    408
         |||||| |||||||                ||||||||        :::
Db 145 TyrSerAlaLeuAlaAsnLysCysHissValGlyCysThrLysArgSerLeuAlaArg    164

Qy 409 TTATGT                                                        414
         |||
Db 165 PheCys                                                        166
```

RESULT 8
A49014
20K protein - rabbit
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 16-Jul-1999
C;Accession: A49014
R;Jetten, A.M.; Bernacki, S.H.; Floyd, E.E.; Saunders, N.A.; Pieniazek, J.; Lotan, R.
Cell Growth Differ. 3, 549-556, 1992
A;Title: Expression of a preprorelaxin-like gene during squamous differentiation of rab
A;Reference number: A49014; MUID:93002619; PMID:1339318
A;Accession: A49014
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-178 <JBT>
A;Cross-references: GB:845940; NID:g257388; PIDN:AAB23648.1; PID:g257389
A;Experimental source: tracheobronchial epithelial cells
A;Note: sequence extracted from NCBI backbone (NCBIN:115816, NCBIP:115821)
C;Superfamily: insulin

Alignment Scores:
Pred. No.:
Score: 0.00144   Length: 178
Percent Similarity: 111.50   Matches: 49
Best Local Similarity: 32.77%   Conservative: 9
Query Match: 27.68%   Mismatches: 60
                        14.41%   Indels: 59
DB: Gaps: 6

US-09-518-842-1 (1-420) x A49014 (1-178)

```
Qy   4 GCCAGCCTGTTCCGGTCCTATCTGCCAGCAATCTGCTCGTGCTGAGCCAA----------  54
         |||||| ||||||  |||||
Db   3 AlaAlaLeuPhe------TyrLeuLeuGlyPheCysLeuLeuGlnGlyGlnValThrGly   20
```

Qy   235 TTAGGTACGACATCAGAATTCATTCCTAATTGTCACCAGAGCTGAGAAAAACCACTGTCT 294
           ::: ||||||| ||||||| |||||| ||||||||||
Db   65   IleIleIleMetLeuGluPheIleAlaAsnLeuProProGluLeuLysAlaAlaLeuSer 84

Qy   295 GAAGGCCAGCCATCATG-- 312
           ||| |||||| |||||
Db   85   GluArgGlnProSerLeuProGlnTyrValProAlaLeuLysAspSerAsn 104

Qy   313 --------------------AAGAAAATAATACTTTCCGCAAAAAGAAGT----------- 345
                               |||:||: :::|||||
Db   105 LeuSerPheGluGluLysLysLeuIleLeuArgAsnArgGlnSerGluAlaAlaAspSer 124

Qy   346 ------------------------------------GGACGTCACAGATTTGATCCA 366
                                               ::: ::: |||||
Db   125 AsnProSerGluLeuLysTyrLeuGlyLeuAspThrHisSerGlnLysLysArgGlnPro 144

Qy   367 TTC---------------------TGTTGAAGTAATTTGTGACGATGGAACTTCAGTTAAA 408
           ::: |||
Db   145 TyrValAlaLeuPheGluLysCysCysLeuIleGlyCysThrLysArgSerLeuAlaAsn 164

Qy   409 TTATGT 414
           |||
Db   165 TyrCys 166

RESULT 10
I47053
relaxin B,C and A chains - horse (fragment)
C;Species: Equus sp.
C;Date: 04-Sep-1997 #sequence_revision 07-Nov-1997 #text_change 16-Jul-1999
C;Accession: I47053
R;Klonisch, T.; Ryan, P.L.; Yamashiro, S.; Porter, D.G.
Biol. Reprod. 52, 1307-1315, 1995
A;Title: Partial complementary deoxyribonucleic acid cloning of equine relaxin messenge
A;Reference number: I47053; MUID:95359320; PMID:7543295
A;Accession: I47053
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-143 <KLO>
A;Cross-references: GB:S78800; NID:g1042059, PIDN:AAB35036.1; PID:g1042060
C;Superfamily: insulin

Alignment Scores:
Pred. No.:   Length: 143
Score: 0.0101   Matches: 30
Percent Similarity: 51.76%   Conservative: 14
Best Local Similarity: 35.29%   Mismatches: 34
Query Match: 13.37%   Indels: 7
DB: Gaps: 3

US-09-518-842-1 (1-420) x I47053 (1-143)

Qy   82   CTGAGGGGATGTGGTCCCGATTTGAAAAACACTTGTCTGTCATATTGCCCATGCCT--- 138
           :: |||| ||| |||||||||| ::: :: |||||| ||||||
Db   1   IleLysAlaCysGlyArgGluLeuAlaArgIleGlulleCysGlySerLeuSer 20

Qy   139 ---GAGAAGACATTC-----ACCACCACCCCAGGAGGGGTGGCTGCTGGAATTCGACGT 189
           :: ||||||| ||||||
Db   21   TrpLysThrValLeuArgGluLeuGlyProGly-------LeuGluValGlyGln 37

Qy   190 CCCAAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAACAAGCCTTAGTACGACATCA 249
           ||| ||||:||||||:::||||||: :||||||: :::||||| :|||||
Db   38   ProValGlyLeuValSerSerSerIleSerLysAspAlaGluAlaLeuAsnThrLysLeu 57

Qy   250 GAATTCATTCCTAATTGTCACCAGAGCTGAGAAAACCACTGTCTGAAGGCCAGCCATCA 309
           ||||||||||||||||||||||||||||||||||| :::
Db   58   GlyLeuAsnSerAsnLeuProLysGluGlnLysAlaThrLeuSerGluArgGlnProSer 77

Qy   310 TTGAAGAAAATAATA 324
           ::::|||||:|:
Db   78   TrpArgGluLeuLeu 82

RESULT 11

Qy   55   ---CTCCTTGAGAGAAGCCTAGCAGCAGCAGACTGAGGGGATGTGGTCCCGATTTGGAAAA 111
           ::: ||||| |||
Db   21   ArgValThrTyrGluTrpMetMetGluAsnVallysIleCysArgAsnAspPheValArg 40

Qy   112 CACTTGCTGTCATATTGC--------------------------CCATGCCTGAGGAG 144
           |||||||
Db   41   ThrAlaIleGluValCysGlyHisValHisLeuGluArgGluSerProSerProGluAsn 60

Qy   145 ACATTCACCACCACCCCAGGAGGGGTGGCTGCTGGAATTCGACGAAATGGTG 204
           ||| ||||| ||||||||
Db   61   ProPhe-----------------------LeuSerSerGlyProAlaAlaGluThrVal 72

Qy   205 TCAACCTCCAACAACAAAGATGGACAAGCCTTAGTACGACATCAGAATTCATTCCTAAT 264
           :::||| ||| ::: || |||||
Db   73   ProSerSerIleLysLysAspAlaGluAsnAlaAsnThrMetLeuGluSerIleProAsn 92

Qy   265 TTGTCACCAGAGCTGAAGAAAACCACTGTCTGAAGGGGCAGCCATCA--------------- 309
           |||||||||||
Db   93   LeuProGlnGluLeuThrArgLeuPheGluLysSerGlnProSerLysLysTyrLeuGln 112

Qy   310 --------------------------------------TTGAAGAAAATAATA--- 324
                                                 ||||||||
Db   113 TyrLeuProThrLeuLysLysSerAsnValSerPheGluGluLysLysIleIleGln 132

Qy   325 ---------------------------------------CTTCCCGCAAA 336
                                                  |||||
Db   133 AsnIleGlnArgGlyValGlnValGlnGlySerSerAlaSerGluSerAsnThrPheSerArgLys 152

Qy   337 AAGAAGAAGTGGACGTCACAGATTTGATCCATTCTGTTGTGAAGTAATTGT 387
           ||||||||:::
Db   153 LysArgGlnPheSerGluSerLeuProGluGluCysCysLysTyrGlyCys 169

RESULT 9
S42783
relaxin 1 precursor - chimpanzee (fragment)
C;Species: Pan troglodytes (chimpanzee)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 16-Jul-1999
C;Accession: S42783
C;Evans, B.B.
A;Submitted to the EMBL Data Library, November 1993
A;Reference number: S42776
A;Accession: S42783
A;Molecule type: mRNA
A;Residues: 1-166 <EVA>
A;Cross-references: EMBL:Z27225; NID:g415996; PIDN:CAA81739.1; PID:g415997
C;Genetics:
C;Gene: rlx1
C;Superfamily: insulin
C;Keywords: disulfide bond; hormone
C;1-5/Domain: signal sequence (fragment) #status predicted <SIG>
C;6-166/Product: relaxin 1 #status predicted <MAT>

Alignment Scores:
Pred. No.:   Length: 166
Score: 0.00299   Matches: 43
Percent Similarity: 37.65%   Conservative: 18
Best Local Similarity: 26.54%   Mismatches: 44
Query Match: 14.02%   Indels: 57
DB: Gaps: 7

US-09-518-842-1 (1-420) x S42783 (1-166)

Qy   61   AGAGAAAGCCTAGCAGCAGCAGACTGAGGGGATGTGGT---CCCGATTTGGAAAACACTTG 117
           ||||| :::|||||||||:::||||||
Db   18   ArgGluLeuValArgAlaGlnIleAlaIleGlyCysGlyMetSerThrTrpSerLysArgSer 37

Qy   118 CTGTCATATTGCCCCATGCCTGAGGAGACATTCACCACCACCCCAGGAGGGGTGGCTGCTG 177
           |||||:::
Db   38   LeuSerGlnGluAspAlaProGln------------ThrPro------------ 47

Qy   178 GAATTCGGACGTCCC--AAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAACAAGCC 234
           |||||||| ::: :::
Db   48   --------ArgProValAlaGluIleValLeuValProSerPheIleAsnLysAspThrGluThr 64

T16833
hypothetical protein T07H6.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Mar-2000
C;Accession: T16833
R;Geisel, C.
submitted to the EMBL Data Library, April 1996
A;Description: The sequence of C. elegans cosmid T07H6.
A;Reference number: Z16586
A;Accession: T16833
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
;Residues: 1-560 <GBI>
;Cross-references: EMBL:U53344; NID:g1255886; PID:g1255889; PIDN:AAA96225.1; GSPDB:GN00
A;Experimental source: strain Bristol N2; clone T07H6
;Genetics:
;Gene: CESP:T07H6.5
;Map position: X
;Introns: 14/1; 75/1; 102/3; 128/1; 186/1; 272/2; 326/1; 361/1; 422/1; 475/1; 527/1
;Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology

Alignment Scores:
Pred. No.:   0.0218   Length: 560
Score: 101.00   Matches: 43
Percent Similarity: 38.56%   Conservative: 16
Best Local Similarity: 28.10%   Mismatches: 38
Query Match: 13.05%   Indels: 56
B: Gaps: 9

US-09-518-842-1 (1-420) x T16833 (1-560)

Qy 94 GGTCCCGATTGGAAAACACTTGCTGTCATATTGCCCCATGCTGAGAAGACATTCACC 153
      ||| ||||| :::||| ||
Db 414 GlyProProMet--LysCysLeuProSerTrpCysGluHisProSerLysThrTyrGly 432

Qy 154 ACCACCCCAGGAGGTGGCTGCTG- 177
      ||| ||||||| :::|||
Db 433 ThrLeuProGlyGlyGlnIleLeuLeuLeuGluGlyIleLeuGlyAlaTyrGluPheGlnSer 452

Qy 178 -----------GAATCTGGACGTCCCAAAGAAATGTGTCAACCTCCAACAACAA 222
      |||| :::|||
Db 453 TyrIleGlnLysValGluGluGlyArg-------AlaIleSerPheGlnCysGlyLys 469

Qy 223 GATGGACAAGCCTAGGT- 264
      :::|||
Db 470 GlyAsnTyrLeuIleGlyProProLysAlaThrCysValAsnGlyGluTrpMetProLys 489

Qy 265 TTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCATTGAAG--AAAATA 321
      ::||||| ||| :::|||:: ||| ||||||
Db 490 ValSerPro-------LysCysValSerGlnThrHisProMetIleGluGlyLysIle 506

Qy 322 ATACTTTCCCGCAAAAAGGAGAAGT-----GGACGT---------- 351
      |||||||||||| :::
Db 507 LeuTrpAspArgLysLysArgSerLeuProGlyArgAlaValArgGluTyrValAspAsp 526

Qy 352 -----------CACAGATTGATCCATTCTGTTGTGAAGTAATT----------- 384
      ||||| :::::
Db 527 GluLeuSerThrHisArgGlnHisSerAspAsnGlyValSerVal 546

Qy 385 -----------TGTGACGATGGAACTTCAGTT 405
      ||:::||| |||||
Db 547 ArgMetIleMetGlnHisSerAspAsnGlyValSerVal 559

RESULT 12
E0614
robable competence-related protein STY0984 [imported] - Salmonella enterica subsp. ente
;Species: Salmonella enterica subsp. enterica serovar Typhi
;Note: this species has also been called Salmonella typhi
;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
;Accession: AE0614
;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
h, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica sero
A;Reference number: AB0S02; MUID:21534947; PMID:11677608
A;Accession: AE0614
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-754 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD05383.1; PID:g16502146; GSPDB:GN00176
C;Genetics:
A;Gene: STY0984

Alignment Scores:
Pred. No.:   0.127   Length: 754
Score: 94.00   Matches: 42
Percent Similarity: 39.13%   Conservative: 21
Best Local Similarity: 26.09%   Mismatches: 54
Query Match: 12.62%   Indels: 45
DB: Gaps: 9

US-09-518-842-1 (1-420) x AE0614 (1-754)

Qy 399 AGTTCCATCGTCACAACAATTACTTCACAACAGAATGATCAAATCTGTGAGTCCACTTCT 340
      ||:::::::::|| |||||
Db 386 ThrAlaMetLeuAlaAsnLeuPheAlaValProTrpValThrPheValThrValProLeu 405

Qy 339 CTTTTGCGGAAAGTAT-----------------TATTTCTTCAATGATGGCTGCCC 298
      ::: ||||| :::|||||||||::: ||||
Db 406 IleLeuAlaGlyMetIleLeuHisLeuThrGlyProPhePheLeuGluGluTrpVal--- 424

Qy 297 TTCAGACAGTGGTTCT----------------- 280
      ||:::|||
Db 425 ----------TrpTyrLeuThrAspArgAlaLeuAlaLeuAlaLeuPheTyrLeuLeuAsnAla 441

Qy 279 ------CAGCTCTGTGACAAATTAGGAATTCTG---ATGTCGTACCTAAGGCTT 230
      ::: :::
Db 442 LeuProGlnGlyTrpValAsnIleAspAsnArgTrpGlnTrpTrpLeuThrLeuProTrp 461

Qy 229 GTCCATCTTGTTGTTGGAGGTGACACCCATTTCTTGGACGTCCAGATTCCAGCAGCC 170
      ::: |||||||||
Db 462 LeuThrLeuIleAlaTrpAsn-----ValTrpArgThrTrp---ProAlaVal 478

Qy 169 ACCCTCCTGGGGTGGTGGTGTCTTCTCCAGGCATGGGCAATATGACGCAAGTGTT 110
      ||||| ::: |||||
Db 479 CysLeuCysGlyLeuLeuLeuMetSerTrpProLeuTrpArgProIleAsnAlaSerGly 498

Qy 109 TTTCCAAATCGGGAGACCACCATCCCCTCAGCTCGTCGTAGGCTTTCTCTAAGGAGTTGGC 50
      :::||:::||| ||| ||||||||
Db 499 TrpGlnVal------HisMetLeuAspVal--------------------GlyGlnGly 509

Qy 49 TCAGCAGCAGCCAGATTGCTGCAGATAGG----------------ACCGGAACAGGCTGG 5
      ||| ||||| :::
Db 510 LeuAlaIleAlaIleValAlaArgGlyAspLysValIleLeuTyrAspThrGlyAlaTrp 529

Qy 4 CCA 2
   |||
Db 530 Pro 530

RESULT 13
E83259
hypothetical protein PA3089 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: E83259
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; P
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic path
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: E83259
A;Status: preliminary
A;Molecule type: DNA

A;Residues: 1-321 <STO>
A;Cross-references: GB:AE004733; GB:AE004091; NID:g9949194; PIDN:AAG06477.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA3089
C;Superfamily: Neisseria meningitidis hypothetical protein NMB1349

Alignment Scores:
Pred. No.:   1.75   Length: 321
Score: 83.00   Matches: 17
Percent Similarity: 61.54%   Conservative: 7
Best Local Similarity: 43.59%   Mismatches: 6
Query Match: 10.72%   Indels: 9
DB: Gaps: 2

US-09-518-842-1 (1-420) x E83259 (1-321)

Qy   94 GGTCCCGATTTGGAAAAACACTTGCTGTCAT----------------ATTGCCCATG 135
         ||||||::: |||||| ||| |||| | ::|||||||
Db   234 GlyProGlnThrGlyGlnHisTrpLeu-HisLeuSerArgGluHisTrpLeuAlaProAl 253

Qy   136 CCTGAGAAGACATTCACCA------CCACCCCAGGAGAGGTGGCTGCTGGAATCTG 184
          ::||||| ||| ||||||||||||||:::||||||
Db   253 aMetArgArgGlnAspProLeuAspProProArgAspGlyAlaCysTrpAsnIle 271

RESULT 14
T09271
probable tail component protein 1626 - Streptococcus thermophilus phage Sfi19
C;Species: Streptococcus thermophilus phage Sfi19
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000
C;Accession: T09271
R;Désiere, F.; Lucchini, S.; Brussow, H.
A;Title: Evolution of Streptococcus thermophilus bacteriophage genomes by modular exchan
A;virology 241, 345-356, 1998
A;Reference number: 216607; MUID:98160788; PMID:9499809
A;Accession: T09271
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1626 <DES>
A;Cross-references: EMBL:AF032122; NID:g2935682; PID:g2935689
A;Keywords: tail protein

Alignment Scores:
red. No.:   2.36   Length: 1626
core: 82.50   Matches: 23
ercent Similarity: 36.25%   Conservative: 6
est Local Similarity: 28.75%   Mismatches: 20
uery Match: 10.66%   Indels: 31
B: Gaps: 3

S-09-518-842-1 (1-420) x T09271 (1-1626)

y   9 CCTGTTCCG------------------------------GTCCTATCTGCC 29
         |||||| ::|||:::|||
b   1074 ProValProGlyArgAsnThrTyrAlaValLeuProAlaAlaGlyAlaValLeuAsnAla 1093

y   30 AGCAATCTGCCTGCTGCTGAGCCAACTCCTTAGAGAAAGCCTAGCAGAGCTGAGGGG 89
         ||| ||||| ::|||||
b   1094 SerGluTrpAlaAlaLeuSerGlyAla----------LysProPheAlaLysGlyThrGly 1110

y   90 ATGTGTGGTCCCCGATTTGGAAAAACACTTGCTGTCATATTGCCCATGCCTGAGAAGACATT 149
         |||||| ||||| |||
b   1111 PheTrpSerLysIleTrpAsnThrThr--------------------------- 1120

y   150 CACCACCACCCCCAGGAGAGGTGGCTGCTGGAATCTGGACGTCCCAAAGAAATGGTGTCAAC 209
                                                ::||| ::|||||
b   1121 ----------AsnValAlaGlyLerValTrpAsnGlyLeuLysAsnGlyValAsp 1135

ESULT 15
02809
robable peptidylprolyl isomerase (EC 5.2.1.8) PPCTIB [similarity] - Leishmania major (B
;Species: Leishmania major
;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 01-Sep-2000

C;Accession: E81457; T02809
R;Myler, P.J.; Audleman, L.; deVos, T.; Hixson, G.; Kiser, P.; Lemley, C.; Magness, C.;
Proc. Natl. Acad. Sci. U.S.A. 96, 2902-2906, 1999
A;Title: Leishmania major Friedlin chromosome 1 has an unusual distribution of protein-
A;Reference number: A81455; MUID:99178987; PMID:10077609
A;Accession: E81457
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-335 <PYL>
A;Cross-references: GB:AE001274; NID:g3264850; PIDN:AAC24632.1; PID:g2995585; GSPDB:GN0
A;Experimental source: strain MHOM/IL/81/Friedlin
C;Genetics:
A;Gene: PPCTIB
A;Map position: 1
C;Superfamily: hypothetical protein PPCTIB; cyclophilin homology
C;Keywords: cis-trans-isomerase
F;149-331/Domain: cyclophilin homology <CYP>

Alignment Scores:
Pred. No.:   2.25   Length: 335
Score: 82.00   Matches: 33
Percent Similarity: 35.94%   Conservative: 13
Best Local Similarity: 25.78%   Mismatches: 28
Query Match: 11.01%   Indels: 54
DB: Gaps: 8

US-09-518-842-1 (1-420) x T02809 (1-335)

Qy   251 TCTGATGTCGTACCTAAGGCTTGTCCATCTTG-----TTGTTGGAGGTTGACACCATT 198
         ::| ||| ||| ||| ::| ||| ||||
Db   172 AlaSerValAlaProLeuThrCysLysAsnPheCysGluLeuCysArgGlyThrThrVal 191

Qy   197 TCTTTGGGACGTCCAGATTCCAGCAGCCAC----------CCTCCT------ 162
         ::||||| ||||| ::|||||||| |||||
Db   192 AlaMetGly------AspThrValAlaHisSerValPheAspProProAlaGln 209

Qy   162 -- 162

Db   210 HisIleGlyTyrLysGlyThrThrPhePheArgThrLeuLysAspAlaTrpValMetGly 229

Qy   161 GGGGTGGTGGTGAATGTCTTCTCAGGCATGGGCAATATGAC-----AGCAAGTGTTTT 108
                            :::||| ||:|:
Db   230 GlyAspValThrGlyAlaHisSerGlyAsnGlyGlyTyrSerCysTyrGlyArgCysPhe 249

Qy   107 CCA---AATCGG 99
         ||| |||
Db   250 ProAspGluThrTyrAlaValProHisAspAlaAlaGlyIleLeuGlyMetCysAsnAsp 269

Qy   98 GGACCACATCCCCTCAGCTCGTCGTGCTCAGGCTTTCCTAAGG------AGTTGGCTCAGC 45
         ||||||||| ::::::||||||
Db   270 GlyProHisThrSerSerSerAlaPheTyrIleThrLeuArgProMetSerTrpMetAsn 289

Qy   44 AGCAGCCAGATTGCT--GGCAGA 24
         ||||||| |||||
Db   290 GlyLysTyrValAlaPheGlyArg 297

Search completed: October 9, 2003, 12:19:46
Job time : 36 secs

GenCore version 5.1.6

Copyright (c) 1993 - 2003 Compugen Ltd.

M nucleic - protein search, using frame_plus_n2p model

un on: October 9, 2003, 12:18:46 ; Search time 51 Seconds
                                      (without alignments)
                                      2653.880 Million cell updates/sec

itle: US-09-518-842-1
ffect score: 774
equence: 1 ATGGCCAGCCTGTTCCGGTC.........CAGTTAAATTATGTACACTAG 420

coring table: BLOSUM62
               Xgapop 10.0 , Xgapext 0.5
               Ygapop 10.0 , Ygapext 0.5
               Fgapop 6.0 , Fgapext 7.0
               Delop   6.0 , Delext   7.0

earched: 600653 seqs, 161128416 residues

otal number of hits satisfying chosen parameters: 1201306

inimum DB seq length: 0
aximum DB seq length: 200000000

ost-processing: Minimum Match 0%
                Maximum Match 100%
                Listing first 45 summaries

ommand line parameters:
MODEL=frame+n2p.model -DEV=xlp
Q=/cgn2_1/USPTO_spool/US09518842/runat_09102003_111040_13189/app_query.fasta_1.583
DB=Published_Applications_AA -QFMT=fastan -SUFFIX=rapb -MINMATCH=0.1
LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
MAXLEN=2000000000 -USER=US09518842 @CGN_1_83 @runat_09102003_111040_13189
NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

atabase :   Published_Applications_AA:*
    1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
    2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
    3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
    4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
    5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
    6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
    7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
    8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
    9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
   10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
   11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
   12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
   13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
   14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
   15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
   16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
   17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
   18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

esult   Query
No.   Score Match Length DB ID   Description
---- ------ ------ ------ --   -----------

No.	Score	Match	Length		Pred.	DB ID	Description
1	228	29.5	41		12	US-10-339-740-223	Sequence 223, App
2	147	19.0	25		12	US-10-339-740-224	Sequence 224, App
3	119.5	15.4	185		15	US-10-205-823-349	Sequence 349, Appli
4	78.5	10.1	890		14	US-10-028-056-3	Sequence 3, Appli
5	76	10.2	619	c	10	US-09-976-059-30	Sequence 30, Appl
6	74.5	10.0	1166	c	15	US-10-101-464A-900	Sequence 900, App
7	74.5	9.6	1531		12	US-09-841-260-98	Sequence 98, Appl
8	74.5	9.6	1531		14	US-10-007-693-98	Sequence 98, Appl
9	74	9.9	857	c	12	US-10-032-585-7857	Sequence 7857, Ap
10	74	9.6	1852		12	US-10-023-649-2	Sequence 2, Appli
11	73.5	9.5	269		9	US-09-815-242-5429	Sequence 5429, Ap
12	73.5	9.5	273		9	US-09-815-242-12487	Sequence 12487, A
13	73.5	9.5	273		9	US-09-815-242-12747	Sequence 12747, A
14	73.5	9.5	274		10	US-09-925-637-12	Sequence 12, Appl
15	73.5	9.5	274		15	US-10-084-205-12	Sequence 12, Appl
16	73.5	9.5	282		12	US-09-772-598-1	Sequence 1, Appli
17	73	9.8	228	c	15	US-10-001-254-22	Sequence 22, Appl
18	73	9.4	458		9	US-09-826-463-1	GENERAL INFORMA
19	73	9.8	560	c	10	US-09-978-295A-420	Sequence 420, App
20	73	9.8	560	c	10	US-09-978-697-420	Sequence 420, App
21	73	9.8	560	c	10	US-09-978-192A-420	Sequence 420, App
22	73	9.8	560	c	10	US-09-999-832A-420	Sequence 420, App
23	73	9.8	560	c	11	US-09-978-189-420	Sequence 420, App
24	73	9.8	560	c	11	US-09-978-608A-420	Sequence 420, App
25	73	9.8	560	c	11	US-09-978-585A-420	Sequence 420, App
26	73	9.8	560	c	11	US-09-978-191A-420	Sequence 420, App
27	73	9.8	560	c	11	US-09-978-403A-420	Sequence 420, App
28	73	9.8	560	c	11	US-09-978-564A-420	Sequence 420, App
29	73	9.8	560	c	11	US-09-999-833A-420	Sequence 420, App
30	73	9.8	560	c	11	US-09-981-915A-420	Sequence 420, App
31	73	9.8	560	c	11	US-09-978-824-420	Sequence 420, App
32	73	9.8	560	c	11	US-09-918-585A-420	Sequence 420, App
33	73	9.8	560	c	11	US-09-978-423A-420	Sequence 420, App
34	73	9.8	560	c	11	US-09-978-193A-420	Sequence 420, App
35	73	9.8	560	c	11	US-09-999-830A-420	Sequence 420, App
36	73	9.8	560	c	11	US-09-978-757B-420	Sequence 420, App
37	73	9.8	560	c	11	US-09-978-187B-420	Sequence 420, App
38	73	9.8	560	c	11	US-09-978-643A-420	Sequence 420, App
39	73	9.8	560	c	12	US-09-978-375A-420	Sequence 420, App
40	73	9.8	560	c	11	US-09-978-188A-420	Sequence 420, App
41	73	9.8	560	c	12	US-09-978-298A-420	Sequence 420, App
42	73	9.8	560	c	12	US-10-143-031A-420	Sequence 420, App
43	73	9.8	560	c	12	US-10-002-967A-420	Sequence 420, App
44	73	9.8	560	c	12	US-10-017-083A-420	Sequence 420, App
45	73	9.8	560	c	12	US-10-143-030A-420	Sequence 420, App

ALIGNMENTS

RESULT 1
US-10-339-740-223

; Sequence 223, Application US/10339740
; Publication No. US20030187246A1
; GENERAL INFORMATION:
;   APPLICANT: Doberstein, Stephen
;   APPLICANT: Reddy, Bindu
;   APPLICANT: Platt, Darren
;   APPLICANT: Ferguson, Kimberly
;   TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS OF C. ELEGANS INSULIN-LIKE GENES AND
;   TITLE OF INVENTION: THEREOF
;   FILE REFERENCE: 7326-069-999
;   CURRENT APPLICATION NUMBER: US/10/339,740
;   CURRENT FILING DATE: 2003-01-09
;   PRIOR APPLICATION NUMBER: US/09/084,303A
;   PRIOR FILING DATE: 1998-05-26
;   NUMBER OF SEQ ID NOS: 298
;   SOFTWARE: PatentIn version 3.1
; SEQ ID NO 223
;   LENGTH: 41
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-339-740-223

```
Alignment Scores:
Pred. No.:                    6.02e-17        Length:           41
Score:                        228.00          Matches:          41
Percent Similarity:           100.00%         Conservative:     0
Best Local Similarity:        100.00%         Mismatches:       0
Query Match:                  29.46%          Indels:           0
DB:                                           Gaps:             0

US-09-518-842-1 (1-420) x US-10-339-740-223 (1-41)

Qy    52 CAACTCCTTAGAGAAAGCCTAGCAGCAGAGCTGAGGGGATGTGGTCCCGATTTGGGAAAA 111
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 GlnLeuLeuArgGluSerLeuAlaAlaGluLeuArgGlyCysGlyProArgPheGlyLys 20

Qy   112 CACTTGCTGTCATATTGCCCCATGCCTGAGAAGACATTCACCACCACCCCAGGAGGGGTG 171
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    21 HisLeuLeuSerTyrCysProMetProGluLysThrPheThrThrThrProGlyGlyTrp 40

Qy   172 CTG 174
         |||
Db    41 Leu 41
```

RESULT 2
US-10-339-740-224
; Sequence 224, Application US/10339740
; Publication No. US2003018724GA1
; GENERAL INFORMATION:
; APPLICANT: Doberstein, Stephen
; APPLICANT: Reddy, Bindu
; APPLICANT: Platt, Darren
; APPLICANT: Ferguson, Kimberly
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS OF C. ELEGANS INSULIN-LIKE GENES AND
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 7326-069-999
; CURRENT APPLICATION NUMBER: US/10/339,740
; CURRENT FILING DATE: 2003-01-09
; PRIOR APPLICATION NUMBER: US/09/084,303A
; PRIOR FILING DATE: 1998-05-26
; NUMBER OF SEQ ID NOS: 298
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 224
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-339-740-224

```
Alignment Scores:
Pred. No.:                    6.53e-08        Length:           25
Score:                        147.00          Matches:          25
Percent Similarity:           100.00%         Conservative:     0
Best Local Similarity:        18.99%          Mismatches:       0
Query Match:                                  Indels:           0
DB:                                           Gaps:             0

US-09-518-842-1 (1-420) x US-10-339-740-224 (1-25)

Qy   343 AGTGGACGTCACAGATTTGATCCATTCTGTTGTGAAGTAATTGTTGTGACGATGGAACTTCA 402
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 SerGlyArgHisArgPheAspProPheCysCysGluValIleCysGluAspGlyThrSer 20

Qy   403 GTTAAAATTATGTACA 417
         |||||||||||||
Db    21 ValLysLeuCysThr 25
```

RESULT 3
US-10-205-823-349
Sequence 349, Application US/10205823
Publication No. US2003010896A1
GENERAL INFORMATION:
APPLICANT: Schlegel, Robert
APPLICANT: Monahan, John E.
APPLICANT: Endege, Wilson O.
APPLICANT: Gannavarapu, Manjula

; APPLICANT: Gorbatcheva, Bella
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Wonsey, Angela M.
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Anderson, Dustin
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF PROSTATE CANCER
; FILE REFERENCE: MRI-044
; CURRENT APPLICATION NUMBER: US/10/205,823
; CURRENT FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 349
; LENGTH: 185
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-205-823-349

```
Alignment Scores:
Pred. No.:                    0.000137        Length:           185
Score:                        119.50          Matches:          52
Percent Similarity:           37.11%          Conservative:     20
Best Local Similarity:        26.80%          Mismatches:       57
Query Match:                  15.44%          Indels:           65
DB:                           15              Gaps:             7

US-09-518-842-1 (1-420) x US-10-205-823-349 (1-185)

Qy     1 ATGGCCAGCCTGCTTCCGTCCGTCCTATCTGCCAGCAATCTGGCTCGCTGCTGAGCCAACTCCTT 60
         ||  ||||||    :::||
Db     1 MetProArgLeuPheLeuPheHisLeuLeuGluPheCysLeuLeuLeuAsnGlnPheSer 20

Qy    61 AGAGAAAGCCTAGCAGCAGAG------------------CTGAGGGGGATGTGGTCCCGATTT 105
         |||   :::|:::||||||::                  :::::   |||||
Db    21 Arg--AlaValAlaAlaLysTrpLysAspAspValIleLysCysGlyArgGluLeu 39

Qy   106 GGAAAACACTTGCTGTCATATTGCCCCATG------------------CCTGAG 141
         ::                                :::
Db    40 ValArgAlaGlyAlaGlnIleAlaIleCysGlyMetSerThrTrpSerLysArgSerLeuSerGln 59

Qy   142 AAGACATTCACCACCACCCCAGGAGGGGTGGCTGCTGGAATCTGGACGTCCC--AAAGAA 198
         :::                                          |||||
Db    60 GluAspAlaProGlnThrPro----------------------------ArgProValAlaGlu 71

Qy   199 ATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGTACGACACATCAGAATTCATT 258
         :::|||||    |||||||  :::   ||||||||||    ::: :::
Db    72 IleValProSerPheIleAsnLysAspThrGluThrIleIleIleMetLeuGluPheIle 91

Qy   259 CCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCATTG-----312
         ||| :::  |||||||||||||||||||||||||||      |||||
Db    92 AlaAsnLeuProProGluLysLysAlaAlaLeuSerGluArgGlnProSerLeuProGlu 111

Qy   313 -------------------------------------------------------AAGAAA 318
                                                                ||||||
Db   112 LeuGlnIleTyrValProAlaLeuLysAspSerAsnLeuserPheGluGluLysByLys 131

Qy   319 ATATACTTTCCGC-------------------------------------------- 333
         :::|||||
Db   132 LeuIleArgAsnArgGlnSerGluAlaAlaAspSerAsnProSerGluLeuLysTyrLeu 151
```

; CURRENT APPLICATION NUMBER: US/09/976,059
; CURRENT FILING DATE: 2001-10-15
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 30
; LENGTH: 619
; TYPE: PRT
; ORGANISM: Actinoplanes sp.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1)
; OTHER INFORMATION: v represents a non-standard initiator codon. It is expected tha
; OTHER INFORMATION: the biosynthesized protein will have a formylmethionine residue
; OTHER INFORMATION: at this position
US-09-976-059-30

Alignment Scores:
Pred. No.:   14.5   Length: 619
Score: 76.00   Matches: 34
Percent Similarity: 39.82%   Conservative: 11
Best Local Similarity: 30.09%   Mismatches: 32
Query Match: 10.20%   Indels: 36
DB: 10   Gaps: 8

US-09-518-842-1 (1-420) x US-09-976-059-30 (1-619)

Qy   253 ATTCTGATGTCGTACCTA------------------AGGCTTGTCCCAT---- 224
           :::|||||||||||||||| ||||:::|||||
Db   214 LeuValThrTyrLeuAlaGlyAlaProValArgAlaArgValAlaHisVal 233

Qy   223 ----------CTTTGTTGTGGAGGTTGACACCATTCTTGGGACGTCCAGATTCCAGCA 173
                     ::: ||||| :::---ValLeuAla 250
Db   234 AlaAlaAlaValAlaAlaThrLeuAlaValSerLeuLeuTrp---------- 250

Qy   172 GCCACCCTCCTGGGTGCGTGGTGAATGTCTTCTCAGGCATGG---------- 131
           |||||||||| :::|||||
Db   251 LeuThrLeuLeuProGly--------SerHisArgProTrpAlaAspThrThrSer 267

Qy   130 GGGCAAATGACAGCAAGTGTTTC----------CAAATCGGGACCACCATC 89
           ||||| ||| ::: ||| :::|||||||
Db   268 GlyAsnAlaPheAlaMetValPheGlyTyrAsnGlyPheAspArgArgAlaGlyIleHisVal 287

Qy   88   CCCTCAGCTCTGCTGCTAGCCTTTCTCTAAGGA----------GTTGGCTCAGCAGCAGCC 38
           ||| ||||||| |||||| ||| :::
Db   288 ProGlyAlaLeuThrArgGlyPheThrArgAspGlyGlyAlaAlaAlaGlyGlySerTrpThr 307

Qy   37   AGATTGCTGGCAGATAGGACCGGAACA------GGCTGG 5
           ||||||||||| |||||
Db   308 AlaLeuAlaAlaAspArgLeuAlaThrArgGlnIleGlyTyrTrp 320

RESULT 6
US-10-101-464A-900
; Sequence 900, Application US/10101464A
; Publication No. US20030046728A1
; GENERAL INFORMATION:
; APPLICANT: Strabala, Timothy
; APPLICANT: Nieuwenhuizen, Nicolaas
; APPLICANT: Higgins, Colleen M.
; TITLE OF INVENTION: Compositions Isolated from Plant Cells
; TITLE OF INVENTION: and Their Use in the Modification of Plant Cell Signaling
; FILE REFERENCE: 11000.1020c2
; CURRENT APPLICATION NUMBER: US/10/101,464A
; CURRENT FILING DATE: 2002-03-18
; PRIOR APPLICATION NUMBER: 09/704,302
; PRIOR APPLICATION NUMBER: 2000-11-01
; PRIOR APPLICATION NUMBER: 09/228,986
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/162,866
; PRIOR FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PCT/US00/00724
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 989
; SOFTWARE: FastSEQ for Windows Version 4.0

Qy   334 --------------------------AAAAGAGAAGTGGACGTCACAGATTTGATCCATTCTGT 372
                                     ||||||||||
Db   152 GlyLeuAspThrHisSerGlnLysLysArgArgProTyrValAlaLeuPheGluLysCys 171

Qy   373 TGTGAAGTAATTGTGACGATGGAACTTCAGTAAATTATGT 414
           ||| ::: |||
Db   172 CysLeuIleGlyCysThrLysArgLeuAlaLysTyrCys 185

RESULT 4
US-10-028-056-3
; Sequence 3, Application US/10028056
; Publication No. US20020152483A1
; GENERAL INFORMATION:
; APPLICANT: REUE, KAREN
; APPLICANT: PTERFY, MIKLOS
; TITLE OF INVENTION: A NOVEL GENE ASSOCIATED WITH REGULATION OF ADIPOSITY AND INSULIN
; FILE REFERENCE: 407T-898010US
; CURRENT APPLICATION NUMBER: US/10/028,056
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: US 60/257,772
; PRIOR FILING DATE: 2000-12-22
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 890
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-028-056-3

Alignment Scores:
Pred. No.:   8.39   Length: 890
Score: 78.50   Matches: 28
Percent Similarity: 40.68%   Conservative: 20
Best Local Similarity: 23.73%   Mismatches: 59
Query Match: 10.14%   Indels: 11
DB: 14   Gaps: 3

US-09-518-842-1 (1-420) x US-10-028-056-3 (1-890)

Qy   34   ATCTGGCTGCTGCTGAGCCAACTCCTTAGAGAAGCCTAGCAGCAGAGCTGAGGGGATGT 93
           ::||| ||| ||| ::: ||
Db   274 LeuTrpGlyGluLeuProGlnAlaAlaLysSerSerSerProHisLysMetLysGluSer 293

Qy   94   GGTCCCCGATTGGAAAA----------CACTTGCTGTCATATTGCCCCATG 135
           ||||||||||||| |||||||| |||||:::
Db   294 SerProLeuSerSerArgLysIleCysAspLysSerHisPheGlnAlaIleHisSerGlu 313

Qy   136 CCTGAGAAGACACATTCACC----------ACCACCCCAGGAGGGTGGCTGCTGGAATCT 183
           ::||||| ||||||| :::
Db   314 SerSerAspThrPheSerAspGlnSerProThrLeuValGlyValGlyValAlaLeuLeuAspGln 333

Qy   184 GGAGCGTCCAAAGAAAATGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACG 243
           ::||||::: ||| ::: |||
Db   334 AsnLysProGlnThrGluMetGlnPheValAsnGluGluGlyLeuGluThrLeuGlyAla 353

Qy   244 ACATCAGAATTCATTCCTAATTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAG 303
           ||||| ::: ||||||| :::
Db   354 AlaAlaProLeuLeuProMetIleGlu--GluLeuLysProSerProSerAlaSerValVal 372

Qy   304 CCATCATTGAAGAAGAAAAATAATACTTTCCCGCAAAAAGAGAAGTGACGTCACAGA 357
           :::||::: ||||||
Db   373 GlnThrAlaAsnLysThrArgAspSerProSerArgGlyArgAspLysArgSerArg 390

RESULT 5
S-09-976-059-30
; Sequence 30, Application US/09976059
; Patent No. US20020164747A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Zazopoulos, Emmanuel
; APPLICANT: Staffa, Alfredo
; TITLE OF INVENTION: Genes and Proteins for Biosynthesis of Ramoplanin.
; FILE REFERENCE: 3019-PCT

; SEQ ID NO 900
; LENGTH: 1166
; TYPE: PRT
; ORGANISM: Pinus radiata
US-10-101-464A-900

Alignment Scores:
Pred. No.:
Score: 25.4
Percent Similarity: 41.30%
Best Local Similarity: 28.26%
Query Match: 10.00%
DB:

Length: 1166
Matches: 39
Conservative: 18
Mismatches: 60
Indels: 21
Gaps: 6

US-09-518-842-1 (1-420) x US-10-101-464A-900 (1-1166)

```
Qy   413 CATAATTAACTGAAGTTCCATCGTCACAAATTACTTCACAACAGAATGATCAAATCTG 354
             ::|||||||||||     |||   :::     ||||||   |||||
Db   164 AsnAsnLeuThrGlyThrIleProSerGlyLeuGlyGlnLeuGlnLeuGlnAsnLeu 183

Qy   353 TGACGTCCACTTCTTCTTTTGCGGAAGTATTATTTC-----TTCAATGATGCTGC 300
             |||:::
Db   184 ------------------------------SerLeuSerPheAsnGlyLeuThrGlyLysVal 194

Qy   299 CCTTCAGACAGTGGTTTCTTCAGCTCTGGTGACAAATTAGGAATGAATTCTGATGTCGTA 240
         |||     |||::: :::   ||||||     |||::: :::
Db   195 ProGluThrPheGlyAsnCysSerAlaLeuGluIleLeuGlyLeuGlySerAsnPheLeu 214

Qy   239 CCTAAGGCTTGTCCATCT------TTGTTGTTGGAGGTGACACCATTTCTTTGGGA 189
         |||||||       :::||
Db   215 SerGlyThrIleProSerHisLeuGlyMetLeuAlaArgIleLysIleLeuSerLeuGly 234

Qy   188 CGTCCAGATTCCAGCAGCCACCCTCCTGGGGTGGTCGTGTGAATGTCTTCTCAGGCATGGGG 129
         |||     |||     |||          :::::|||
Db   235 ArgAsnGlnLeuSerGlyGlyIleProProSerLeuLeuAsnCysThrGluLeuArgGly 254

Qy   128 CAATATGACAGCAAGTGTTTTCCAAATCGG-----GGACCACATCCCCCAGCCTCTGCT 75
         |||     |||||||     |||     |||          |||
Db   255 LeuTyr------------LeuTyrGluAsnArgLeuThrGlyGluIleProTrpGluIleGly 271

Qy    74 GCTAGGCTTTCTCTAAGGAGTGGCTCAGC-----AGCAGCCAGATTGCTGGC 27
         |||:::||||||     |||     ::||:::|||:::|||
Db   272 AlaLysLeuSerLysIleGluThrLeuSerLeuGlyGlyAsnGlnLeuSerGly 289
```

RESULT 7
US-09-841-260-98
; Sequence 98, Application US/09841260
; Publication No. US2003017570A1
GENERAL INFORMATION:
  APPLICANT: Bhatia, Ajay
  APPLICANT: Probst, Peter
  APPLICANT: Stromberg, Erika Jean
  TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND DIAGNOSIS
  TITLE OF INVENTION: OF CHLAMYDIAL INFECTION
  FILE REFERENCE: 210121.515
  CURRENT APPLICATION NUMBER: US/09/841,260
  CURRENT FILING DATE: 2001-04-23
  NUMBER OF SEQ ID NOS: 140
  SEQ ID NO 98
  LENGTH: 1531
  TYPE: PRT
  ORGANISM: Chlamydia trachomatis serovar D
US-09-841-260-98

Alignment Scores:
Pred. No.:
Score: 27.3
Percent Similarity: 44.50%
Best Local Similarity: 30.21%
Query Match: 9.63%
DB:

Length: 1531
Matches: 29
Conservative: 14
Mismatches: 26
Indels: 27
Gaps: 5

US-09-518-842-1 (1-420) x US-09-841-260-98 (1-1531)

```
Qy    79 GAGCTGAGGGGATGTGGTCCCGATTTGGAAAACACTTG--------- 117
         |||     :::|||   |||     |||:::::::
Db   472 GluTyrGlnGlyGlyGlyTyrGlyAlaLeuPheGlyGluAsnIleSerLeuSerGluAsnAlaGly 491

Qy   118 ---CTGTCATATTGCCCATGCCTGAGAAGACATTCACCACCACC-----CCA 162
            ||||:::::::                |||||||
Db   492 ValLeuThrPheLysAspAsnIleValLysThrPheAlaSerAsnGlyLysIleLeuGly 511

Qy   163 GGAGGGTGGCTGCTGGAATCTGACGTCCCAAAGAAATGGTGTCAACCTCCAACAACAAA 222
         |||||||     :::|||   :::|||||      |||::||||
Db   512 GlyGlyAlaIleLeuAlaThrGlyLys--------------ValGluIleThrAsnAsnSer 527

Qy   223 GATGGA---------                                CAAGCCTTAGTGTACGACATCAGAA 252
         :::|||                                         |||||||||   |||   |||
Db   528 GluGlyIleSerPheThrGlyAsnAlaArgAlaProGlnAlaLeuProThrGlnGluGlu 547

Qy   253 TTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGG 300
         |||   |||   ::   :::|||:::||||||||
Db   548 Phe---ProLeuPheSerLysGlyLysGluGlyArgProLeuSerSerGly 562
```

RESULT 8
US-10-007-693-98
; Sequence 98, Application US/10007693
; Publication No. US2002146776A1
GENERAL INFORMATION:
  APPLICANT: Bhatia, Ajay
  APPLICANT: Probst, Peter
  TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT
  TITLE OF INVENTION: AND DIAGNOSIS OF CHLAMYDIAL INFECTION
  FILE REFERENCE: 210121.515C2
  CURRENT APPLICATION NUMBER: US/10/007,693
  CURRENT FILING DATE: 2001-12-05
  NUMBER OF SEQ ID NOS: 157
  SEQ ID NO 98
  LENGTH: 1531
  TYPE: PRT
  ORGANISM: Chlamydia trachomatis serovar D
US-10-007-693-98

Alignment Scores:
Pred. No.:
Score: 27.3
Percent Similarity: 44.50%
Best Local Similarity: 30.21%
Query Match: 9.63%
DB:

Length: 1531
Matches: 29
Conservative: 14
Mismatches: 26
Indels: 27
Gaps: 5

US-09-518-842-1 (1-420) x US-10-007-693-98 (1-1531)

```
Qy    79 GAGCTGAGGGGATGTGGTCCCGATTTGGAAAACACTTG--------- 117
         |||     :::|||   |||     |||:::::::
Db   472 GluTyrGlnGlyGlyGlyTyrGlyAlaLeuPheGlyGluAsnIleSerLeuSerGluAsnAlaGly 491

Qy   118 ---CTGTCATATTGCCCATGCCTGAGAAGACATTCACCACCACC-----CCA 162
            ||||:::::::                |||:::
Db   492 ValLeuThrPheLysAspAsnIleValLysThrPheAlaSerAsnGlyLysIleLeuGly 511

Qy   163 GGAGGGTGGCTGCTGGAATCTGACGTCCCAAAGAAATGGTGTCAACCTCCAACAACAAA 222
         |||||||     :::|||   :::|||||      |||::||||
Db   512 GlyGlyAlaIleLeuAlaThrGlyLys--------------ValGluIleThrAsnAsnSer 527

Qy   223 GATGGA---------                                CAAGCCTTAGTGTACGACATCAGAA 252
         :::|||                                         |||||||||   |||   |||
Db   528 GluGlyIleSerPheThrGlyAsnAlaArgAlaProGlnAlaLeuProThrGlnGluGlu 547

Qy   253 TTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGG 300
         |||   |||   ::   :::|||:::||||||||
Db   548 Phe---ProLeuPheSerLysGlyLysGluGlyArgProLeuSerSerGly 562
```

RESULT 9
US-10-032-585-7857
; Sequence 7857, Application US/10032585

Publication No. US20030180953A1
GENERAL INFORMATION:
APPLICANT: Terry, Roemer D.
APPLICANT: Bo, Jiang
APPLICANT: Charles, Boone
APPLICANT: Howard, Bussey
TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
FILE REFERENCE: 10182-005-999
CURRENT APPLICATION NUMBER: US/10/032,585
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 8000
SOFTWARE: PatentIn version 3.1
SEQ ID NO 7857
LENGTH: 857
TYPE: PRT
ORGANISM: Candida albicans
US-10-032-585-7857

Alignment Scores:
Pred. No.:   26.6   Length: 857
Score: 74.00   Matches: 31
Percent Similarity: 40.74%   Conservative: 13
Best Local Similarity: 28.70%   Mismatches: 44
Query Match: 9.93%   Indels: 20
DB: 12   Gaps: 4

US-09-518-842-1 (1-420) x US-10-032-585-7857 (1-857)

Qy   371 CAGAATGGATCAAATCTGTGACGTCCACTTCTCTTTTCGCGGAAAGTATTATTTTCTTC 312
             ::::|||::||||| ||| :::
Db   443 GluAsnAspAlaAsnLeu------ProTrpTyrPheSerLys------ 454

Qy   311 AATGATGGCTGCCCTTCAGACAGTGGTTTCTTCAGCTCTCGTGACAAATTA-------- 261
             ||||||| :::||| ||||
Db   455 ------------SerAspGlySerPheAlaLysAsnGlyGlnIleLeuPheTrpTyr 469

Qy   260 GGAATGAATTCTGATGTCGTACCTAAGGCTTGTCCATCTTGTTGTTGGAGGTTGACACC 201
             |||:::||| ||| ||| |||:::
Db   470 GlyLeuAsnIleAspGluAlaSerLys---------LeuIleLysGluPheAspSer 485

Qy   200 ATTTCTTTGGGACGTCCAGATTCCAGCAGCCACCCTCCTGGGGTGGTGGGTGAATGTCTTC 141
             ||||||| ||||||
Db   486 SerSerIleGlySerSerLeuSerSerSerGluserGlyValPheThrSerAlaGln 505

Qy   140 TCAGGCATGGGGCAATATGACAGCAAGTGTTTCCAAATCGGGGACCACATCCCCTCAGC 81
             ||| ::: ||| |||:::|||
Db   506 GlnLysArgGlyPheHisHisSerThrValArgArgAsnThrAsnProAsnProProLeu 525

Qy   80 TCTGCTGCTAGGCTTTCTCTCTAAGG 57
             ||| ::: |||
Db   526 SerGluGlyLysGlnThrGluArg 533

RESULT 10
US-10-023-649-2
Sequence 2, Application US/10023649
Publication No. US20030143201A1
GENERAL INFORMATION:
APPLICANT: The Minister of National Defence, Government of Canada
APPLICANT: Nagata, Leslie P
APPLICANT: Wong, Jonathon P
TITLE OF INVENTION: No. US20030143201A1el DNA-Based Vaccine Against the Encephalitis
FILE REFERENCE: NEL-001
CURRENT APPLICATION NUMBER: US/10/023,649
CURRENT FILING DATE: 2002-12-27
PRIOR APPLICATION NUMBER: 60/256,948
PRIOR FILING DATE: 2000-12-21
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 1852
TYPE: PRT
ORGANISM: Western equine encephalomyelitis virus - strain 71V-1658
US-10-023-649-2

Alignment Scores:
Pred. No.:   32.7   Length: 1852
Score: 74.00   Matches: 32
Percent Similarity: 40.69%   Conservative: 27
Best Local Similarity: 22.07%   Mismatches: 40
Query Match: 9.56%   Indels: 46
DB: 12   Gaps: 6

US-09-518-842-1 (1-420) x US-10-023-649-2 (1-1852)

Qy   4 GCCAGCCTGTTCCGGTCCTATCTGCCAGCAATCTGGCTGCTGCTGAGCCAACTCCTTAGA 63
             ||| :::||| ::||| :::|||
Db   501 AlaGluGluIleArgGluAlaLeuProPro------LeuLeuProGluIleGluLys 517

Qy   64 GAAAGCCTAGCAGCAGCAGAGCTG-----------------AGGGGATGTGGTCCCCGATT 105
             |||:::::|||::||||| |||
Db   518 GluThrValGluAlaGluValAspLeuIleMetGlnGluAlaGlyAlaGlySer------ 535

Qy   106 GGAAAACACTTGCTGTCATATTGCCCCATGGCCTGAGAAGACATTCACCACCACCACCAGGA 165
             ||| :::
Db   536 ------------------------------------ValGluThrProArg 540

Qy   166 GGGTGGCTGCTGGAATCTGACGTCCCAAAGAAATGGTGTCAACCTCCAACAACAAAGAT 225
             ||| ::: ||| |||
Db   541 GlyHisIleArgValThrSerTyrProGlyGlu----- 551

Qy   226 GGACAAGCCTTAGGTACGACACATCAGAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAA 285
             :::   :::||| ::: :::   :::|||
Db   552 ---GluLysIleGlySerTyrAlaIleLeuSerProGlnAlaValLeuAsnSerGluLys 570

Qy   286 --------------CCACTGTCTGAAGGGCAGCCATCATTGAAGAAAATAATACTTTCC 330
                           :::|||
Db   571 LeuAlaCysIleHisProLeuAlaGlu----------------GlnValLeuValMetThr 585

Qy   331 CGCAAAAGAGAAGTGGACGTCACAGATTGATCCATTCGTGTTGTGAAGTAATTTGTGAC 390
             ||| :::|||:::|||
Db   586 HisLysGlyArgAlaGlyArgTyrLysValGluProTyrHisGlyLysValIleValPro 605

Qy   391 GATGGAACTTCAGTT 405
             :::|||||:::|||
Db   606 GluGlyThrAlaVal 610

RESULT 11
US-09-815-242-5429
Sequence 5429, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 5429
  LENGTH: 269
  TYPE: PRT
  ORGANISM: Staphylococcus aureus
US-09-815-242-5429

Alignment Scores:
Pred. No.:   22.1
Score: 73.50
Percent Similarity: 44.23%
Best Local Similarity: 28.85%
Query Match: 9.50%
DB: 9

Length: 269
Matches: 30
Conservative: 16
Mismatches: 30
Indels: 28
Gaps: 5

US-09-518-842-1 (1-420) x US-09-815-242-5429 (1-269)

Qy   76 GCAGAGCTGAGGGGATGTGGTCCCCGATTGGAAAAACACTTGCTGTCATATTGCCCCATG 135
          ||| ::: ||||| ||||:: |||||||: ||||||:||
Db   178 AlaProIlePheGlyLeuAsnLysArgGlnGlyArgGlnLeuLeuAlaTyrLeuGlyAla 197

Qy   136 CCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCCAAA 195
          |||:|::: ::: |||||| ::: |||||:::
Db   198 ProLysGluLeuTyrGluLysThrProThrAla--AspLeuGluAspAspLysProGln 216

Qy   196 GAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACACATCAGAATTC 255
                                            ||| |||||||
Db   217 ----------------------------------LeuProAspGluAspAlaLeuGlyValThrTyrGluAla 229

Qy   256 ATTCCTAAT------------------------TTGTCACCAGAG-------------CTGAAGAAA 285
          ||| ||| :::::::|||||
Db   230 IleAspAsnTyrLeuGluGlyLysProValThrProGluGlnLysValIleGlu-As 249

Qy   286 CCACTGTCTGAAGGGCAGCCATCATTGAAGAAAATAATACTTTCCCGCAAAAAGAGAAGT 345
          |||| ::::: ::||| ||||||:::
Db   249 nHisTyrIleArgAsnAlaHis----------------------------LysArgGluLe 260

Qy   346 GGACGTCACA 355
          |||
Db   260 uAlaTyrThr 263

RESULT 12
US-09-815-242-12487
Sequence 12487, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 12487
  LENGTH: 273
  TYPE: PRT
  ORGANISM: Staphylococcus aureus
US-09-815-242-12487

Alignment Scores:
Pred. No.:   22.2
Score: 73.50
Percent Similarity: 44.23%
Best Local Similarity: 28.85%
Query Match: 9.50%
DB: 9

Length: 273
Matches: 30
Conservative: 16
Mismatches: 30
Indels: 28
Gaps: 5

US-09-518-842-1 (1-420) x US-09-815-242-12487 (1-273)

Qy   76 GCAGAGCTGAGGGGATGTGGTCCCCGATTGGAAAAACACTTGCTGTCATATTGCCCCATG 135
          ||| ::: ||||| ||||:: |||||||: ||||:||
Db   181 AlaProIlePheGlyLeuAsnLysArgGlnGlyArgGlnLeuLeuAlaTyrLeuGlyAla 200

Qy   136 CCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCCAAA 195
          |||:|::: ::: ||||| ::: |||||:::
Db   201 ProLysGluLeuTyrGluLysThrProThrAla--AspLeuGluAspAspLysProGln 219

Qy   196 GAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACACATCAGAATTC 255
                                            ||| |||||||
Db   220 ----------------------------------LeuProAspGluAspAlaLeuGlyValThrTyrGluAla 232

Qy   256 ATTCCTAAT------------------------TTGTCACCAGAG-------------CTGAAGAAA 285
          ||| ||| :::::::|||||
Db   233 IleAspAsnTyrLeuGluGlyLysProValThrProGluGlnLysValIleGlu-As 252

Qy   286 CCACTGTCTGAAGGGCAGCCATCATTGAAGAAAATAATACTTTCCCGCAAAAAGAGAAGT 345
          |||| ::::: ::||| ||||||:::
Db   252 nHisTyrIleArgAsnAlaHis----------------------------LysArgGluLe 263

Qy   346 GGACGTCACA 355
          :||
Db   263 uAlaTyrThr 266

RESULT 13
US-09-815-242-12747
Sequence 12747, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 12747
LENGTH: 273
TYPE: PRT
ORGANISM: Staphylococcus aureus

US-09-815-242-12747

Alignment Scores:
Pred. No.:　　　　　　　　22.2　　　　Length:　　　　　273
Score:　　　　　　　　　　73.50　　　　Matches:　　　　30
Percent Similarity:　　44.23%　　　Conservative:　16
Best Local Similarity:　28.85%　　Mismatches:　　30
Query Match:　　　　　　　9.50%　　　　Indels:　　　　　28
DB:　　　　　　　　　　　　　　　　　　Gaps:　　　　　　9

US-09-518-842-1 (1-420) x US-09-815-242-12747 (1-273)

Qy　　76　GCAGAGCTGAGGGGATGTGGTCCCGATTTGGAAAAACACTTGCTGTCATATTGCCCCATG　135
　　　　　　||| :: ||| ||| ||:::|||
Db　181　AlaProIlePheGlyLeuAsnLysArgGlnGlyArgGlnLeuLeuAlaTyrLeuGlyAla　200

Qy　136　CCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCCAAA　195
　　　　　　|||:::: ||| ::: :: |||||
Db　201　ProLysGluLeuTyrGluLysThrProThrAla---AspLeuGluAspLysProGln　219

Qy　196　GAAATGGTGTCAACCTCCAACAACAAGATGGACAAGCCTTAGGTACGACATCAGAATTC　255
　　　　　　||||||| ||| |||
Db　220　---------------------LeuProAspGluAspAlaLeuGlyValThrTyrGluAla　232

Qy　256　ATTCCTAAT------------------TTGTCACCAGAG---------------CTGAAGAAA　285
　　　　　　||| |||　　　:::::||||||
Db　233　IleAspAsnTyrLeuGlyLysProValThrProGluGlnLysValIleGlu-As　252

Qy　286　CCACTGTCTGAAGGGCAGCAGCCATCATTGAAGAAAAATAATACTTTCCCGCAAAAAGAGAAGT　345
　　　　　　|||　　||||||　：：||||
Db　252　nHisTyrIleArgAsnAlaHis----------------------------LysArgGluGlu　263

Qy　346　GGACGTCACA　355
　　　　　　|||
Db　263　uAlaTyrThr　266

ESULT 14
S-09-925-637-12
　Sequence 12, Application US/09925637
　Patent No. US2002010333BA1
　GENERAL INFORMATION:
APPLICANT: Choi
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Polypeptides
FILE REFERENCE: PB560
CURRENT APPLICATION NUMBER: US/09/925,637
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/23773
PRIOR FILING DATE: 2000-08-31
PRIOR APPLICATION NUMBER: US 60/151,933
PRIOR FILING DATE: 1999-09-01
PRIOR APPLICATION NUMBER: US 08/781,986
PRIOR FILING DATE: 1997-01-03
PRIOR APPLICATION NUMBER: US 08/956,171
PRIOR FILING DATE: 1997-10-20
PRIOR APPLICATION NUMBER: US 60/009,861
PRIOR FILING DATE: 1996-01-06
NUMBER OF SEQ ID NOS: 74
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 12
LENGTH: 274
TYPE: PRT
ORGANISM: Homo sapiens

S-09-925-637-12

Alignment Scores:
Pred. No.:　　　　　　　　22.2　　　　Length:　　　　　274
Score:　　　　　　　　　　73.50　　　　Matches:　　　　30
Percent Similarity:　　44.23%　　　Conservative:　16
Best Local Similarity:　28.85%　　Mismatches:　　30
Query Match:　　　　　　　9.50%　　　　Indels:　　　　　28
DB:　　　　　　　　　　10　　　　　　Gaps:　　　　　　5

US-09-518-842-1 (1-420) x US-09-925-637-12 (1-274)

Qy　　76　GCAGAGCTGAGGGGATGTGGTCCCGATTTGGAAAAACACTTGCTGTCATATTGCCCCATG　135
　　　　　　||| :: ||| ||| ||:::|||
Db　182　AlaProIlePheGlyLeuAsnLysArgGlnGlyArgGlnLeuLeuAlaTyrLeuGlyAla　201

Qy　136　CCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCCAAA　195
　　　　　　|||:::: ||| ::: :: |||||
Db　202　ProLysGluLeuTyrGluLysThrProThrAla---AspLeuGluAspLysProGln　220

Qy　196　GAAATGGTGTCAACCTCCAACAACAAGATGGACAAGCCTTAGGTACGACATCAGAATTC　255
　　　　　　||||||| ||| |||
Db　221　---------------------LeuProAspGluAspAlaLeuGlyValThrTyrGluAla　233

Qy　256　ATTCCTAAT------------------TTGTCACCAGAG---------------CTGAAGAAA　285
　　　　　　||| |||　　　:::::||||||
Db　234　IleAspAsnTyrLeuGlyLysProValThrProGluGlnLysValIleGlu-As　253

Qy　286　CCACTGTCTGAAGGGCAGCAGCCATCATTGAAGAAAAATAATACTTTCCCGCAAAAAGAAGT　345
　　　　　　|||　　||||||　::||||
Db　253　nHisTyrIleArgAsnAlaHis----------------------------LysArgGluE　264

Qy　346　GGACGTCACA　355
　　　　　　|||
Db　264　uAlaTyrThr　267

RESULT 15
US-10-084-205-12
　Sequence 12, Application US/10084205
　Publication No. US20030049648A1
　GENERAL INFORMATION:
APPLICANT: Choi, Gil
TITLE OF INVENTION: 37 Staphylococcus aureus Genes and Polypeptides
FILE REFERENCE: PB515P1
CURRENT APPLICATION NUMBER: US/10/084,205
CURRENT FILING DATE: 2002-02-28
PRIOR APPLICATION NUMBER: PCT/US00/23773
PRIOR FILING DATE: 2000-08-31
PRIOR APPLICATION NUMBER: 60/151,933
PRIOR FILING DATE: 1999-09-01
NUMBER OF SEQ ID NOS: 74
SOFTWARE: PatentIn Ver. 3.1
SEQ ID NO 12
LENGTH: 274
TYPE: PRT
ORGANISM: Staphylococcus aureus

US-10-084-205-12

Alignment Scores:
Pred. No.:　　　　　　　　22.2　　　　Length:　　　　　274
Score:　　　　　　　　　　73.50　　　　Matches:　　　　30
Percent Similarity:　　44.23%　　　Conservative:　16
Best Local Similarity:　28.85%　　Mismatches:　　30
Query Match:　　　　　　　9.50%　　　　Indels:　　　　　28
DB:　　　　　　　　　　15　　　　　　Gaps:　　　　　　5

US-09-518-842-1 (1-420) x US-10-084-205-12 (1-274)

Qy　　76　GCAGAGCTGAGGGGATGTGGTCCCGATTTGGAAAAACACTTGCTGTCATATTGCCCCATG　135
　　　　　　||| :: ||| ||| ||:::|||
Db　182　AlaProIlePheGlyLeuAsnLysArgGlnGlyArgGlnLeuLeuAlaTyrLeuGlyAla　201

Qy　136　CCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCCAAA　195
　　　　　　|||:::: ||| ::: :: |||||

```
Db   202 ProLysGluLeuTyrGluLysThrProThrAla---AspLeuGluAspAspLysProGln 220

Qy   196 GAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGTACGACATCAGAATTC 255
             |||      ||||||||||  ||| ||| |||
Db   221 ----------------LeuProAspGluAspAlaLeuGluGlyValThrTyrGluAla 233

Qy   256 ATTCCTAAT----------------TTGTCACCAGAG----------CTGAAGAAA 285
             |||   |||        :::::||||||        :::::: ||
Db   234 IleAspAsnTyrLeuGluGlyLysProValThrProGluGluGlnLysValIleGlu-As 253

Qy   286 CCACTGTCTGAAGGGCAGCCATCATTGAAGAAATAATACTTCCCGCAAAAAGGAGAAGT 345
             ||||  :::::::  :::|||                    ||||||||||::
Db   253 nHisTyrIleArgGlnAsnAlaHis-------------------LysArgGluLeu 264

Qy   346 GGACGTCACA 355
             |  ||
Db   264 uAlaTyrThr 267
```

Search completed: October 9, 2003, 12:32:15
Job time : 55 secs

DM nucleic - protein search, using frame_plus_n2p model

Copyright (c) 1993 - 2003 Compugen Ltd.

Run on:　　　October　9, 2003, 12:08:36 ; Search time 61.5 Seconds
　　　　　　　　　　　　　　　　　　　　(without alignments)
　　　　　　　　　　　　　　　　　　　　2167.973 Million cell updates/sec

Title:　　　　　US-09-518-842-1
Perfect score:　774
Sequence:　　1 ATGGCCAGCCTGTTCCGGTC.........CAGTTAAATTATGTACATAG 420

Scoring table:　BLOSUM62
　　　　　　　　Xgapop 10.0 , Xgapext　0.5
　　　　　　　　Ygapop 10.0 , Ygapext　0.5
　　　　　　　　Fgapop　6.0 , Fgapext　7.0
　　　　　　　　Delop　6.0 , Delext　7.0

Searched:　　1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters:　　　2215726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
　　　　　　　　Maximum Match 100%
　　　　　　　　Listing first 45 summaries

Command line parameters:
MODEL=frame+ n2p.model -DEV=xlp
Q=/cgn2_1/USPTO_spool/US09518842/runat_09102003_111035_13009/app.query.fasta_1.583
DB=A Geneseq_19Jun03 -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
LOOPEXT=0 -UNITS=bits　-START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
USER=US09518842 @CGN_1_137 @runat_09102003_111035_13009 -NCPU=6 -ICPU=3
NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -THREADS=1 -XGAPEXT=0.5 -FGAPOP=6
FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DBLOP=6 -DELEXT=7

atabase :　A Geneseq_19Jun03:*
　1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
　2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
　3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
　4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
　5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
　6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
　7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
　8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
　9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	739	95.5	139	17	AAR89134	Human early placen
2	739	95.5	139	18	AAW17676	Human relaxin-rela
3	739	95.5	139	19	AAW69168	Zins1 protein. Ho
4	739	95.5	139	20	AAY26926	Human insulin-like
5	739	95.5	139	20	AAW99574	Human early placen
6	620	80.1	124	19	AAW69169	Zins1 NF protein.
7	619	80.0	159	19	AAW69170	N-terminally tagge
8	180	23.3	182	4	AAP30392	Sequence of porcin
9	173	22.4	182	5	AAP40156	Sequence of porcin
10	123.5	16.0	185	10	AAP94621	Amino acid sequenc
11	120.5	15.6	185	5	AAP40108	Sequence of human
12	119.5	15.5	185	5	AAP40154	Sequence of human
13	119.5	15.4	185	5	AAP40155	Sequence of human
14	119.5	15.4	185	23	ABG61819	Prostate cancer-as
15	113.5	14.7	150	16	AAR64904	Prolrelaxin fragme
16	113.5	14.7	162	11	AAR07987	H2 prorelaxin gene
17	112.5	14.5	185	10	AAP94622	Amino acid sequenc
18	111.5	14.4	178	14	AAR31958	Squamous cell spec
19	111	14.3	220	22	AAU02911	Angiotensin conver
C 20	103	13.8	787	22	ABG05453	Novel human diagno
C 21	103	13.8	787	22	ABG08971	Novel human diagno
22	101.5	13.1	164	11	AAR07988	H2 prorelaxin dedu
23	88.5	11.4	754	22	AAB93604	Human protein sequ
24	88.5	11.4	754	23	ABB97226	Novel human protei
25	80	10.3	239	21	AAY70362	Human insulin-like
26	80	10.3	240	21	AAY93598	A human insulin-li
27	80	10.3	322	13	AAR22209	h-IGFBP-4. Homo s
28	80	10.3	324	13	AAR21833	Sequence of insuli
C 29	78.5	10.1	912	24	ABP75533	Human secretory po
C 30	78	10.1	210	22	AAU49608	Propionibacterium
31	78	10.1	927	22	ABG25847	Novel human diagno
C 32	76.5	10.3	402	21	AAG29442	Arabidopsis thalia
C 33	76.5	9.9	468	22	AAU52145	Propionibacterium
C 34	76	10.2	212	24	ABJ18429	Breast specific re
C 35	76	10.2	619	23	AAO22174	Ramoplanin biosynt
C 36	76	10.2	740	22	AAM93401	Human polypeptide,
37	76	9.8	812	24	ABU69602	Human NF-kappaB as
38	75.5	9.8	612	22	ABG11791	Novel human diagno
39	75.5	9.8	1034	22	ABG21134	Novel human diagno
C 40	75	10.1	994	18	AAW27147	HMG-CoA reductase
C 41	75	10.1	1711	19	AAW70506	Osteotesticular pr
C 42	75	10.1	1711	19	AAW70507	Mutant osteotestic
C 43	74.5	10.0	135	23	ABP34075	Human ORF3048 prot
C 44	74.5	10.0	1166	21	AAB25525	Pinus radiata cell
C 45	74.5	9.6	1531	22	AAU38904	C. trachomatis CT8

ALIGNMENTS

RESULT 1
AAR89134
ID　AAR89134 standard; Protein; 139 AA.
XX
AC　AAR89134;
XX
DT　31-JUL-1996　(first entry)
XX
DE　Human early placental insulin-like protein.
XX
KW　Insulin growth hormone family; early placental insulin like protein;
KW　placentin; placental tissue; cytotrophoblast; trimester; probe; primer;
KW　amplification; polymerase chain reaction; tyrosine phosphorylation;
KW　cellular protein; growth factor; human; lactation; promoter; PCR;
KW　regeneration; nerve; muscle; skin; bone tissue.
XX
OS　Homo sapiens.
XX

```
Qy    241  ACGACATCAGAATTCATTCCTAATTGTCACCAGAGCTGAAGAAACCACTGTTGTTGAAGGG 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81  ThrThrSerGluPheIleProAsnLeuSerProGluLeuLysProLeuSerGluGly 100

Qy    301  CAGCCATCATTGAAGAAAATATACTTTCCCGCAAAAAGAGAAGTGGACGTCACAGATTT 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    101  GlnProSerLeuLysLysIleIleLeuSerArgLysLysArgSerGlyArgHisArgPhe 120

Qy    361  GATCCATTCTGTTGTGAAGTAATTGTGACGATGGAACTTCAGTTAAATTATGTACA 417
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  AspProPheCysCysGluValIleCysCysAspAspGlyThrSerValLysLeuCysThr 139
```

RESULT 2
AAW17676
ID   AAW17676 standard; Protein; 139 AA.
XX
AC   AAW17676;
XX
DT   24-JUL-1997 (first entry)
XX
DE   Human relaxin-related factor-2 (RRF-2).
XX
KW   Relaxin-related factor-2; RRF-2; testis; sperm; growth factor;
KW   infertility.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT
FT   Peptide   1..26
FT   /label= Sig_peptide
FT   Misc-difference 124
FT   /note= "conserved Cys residue indicative of
FT   insulin family member"
FT   Misc-difference 125
FT   /note= "conserved Cys residue indicative of
FT   insulin family member"
FT   Misc-difference 129
FT   /note= "conserved Cys residue indicative of
FT   insulin family member"
FT   Misc-difference 138
FT   /note= "conserved Cys residue indicative of
FT   insulin family member"
XX
PN   WO9716549-A2.
XX
PD   09-MAY-1997.
XX
PF   01-NOV-1996;   96WO-US17342.
XX
PR   21-FEB-1996;   96US-0012016.
PR   03-NOV-1995;   95US-0006221.
XX
PA   (REGE-) REGENERON PHARM INC.
XX
PI   Davis S;
XX
DR   WPI; 1997-272118/24.
DR   N-PSDB; AAT68419.
XX
PT   New isolated relaxin-related factor genes - used to develop products
PT   which can be used in diagnosis and therapy, e.g. in fertility and
PT   pregnancy applications
XX
PS   Example 3; Fig 4; 34pp; English.
XX
CC   Human relaxin-related factor-2 (RRF-2) (AAW17676) is a placenta-
CC   specific growth factor related to relaxin and to the insulin family
CC   of ligands. RRF-2 cDNA (AAT68419) was isolated in a search of
CC   expressed sequence tags for sequences related to relaxin. RRF-2
CC   displays all the expected features of a new insulin family member,
CC   partic. with regard to a cluster of four cysteine residues at the
CC   C-terminus of the molecule. A related testis-specific factor,
CC   RRF-1 (AAW17675), has also been identified. RRF-2 can be produced

---

FH   Key   Location/Qualifiers
FT   Peptide   1..17
FT   /note= "signal peptide"
FT   Peptide   18..58
FT   /note= "B-chain peptide"
FT   Peptide   59..109
FT   /note= "C-peptide, links B and A chains in pre-E"
FT   Peptide   110..139
FT   /note= "A-chain peptide"
XX
PN   WO9534653-A1.
XX
PD   21-DEC-1995.
XX
PF   12-JUN-1995;   95WO-FR00766.
XX
PR   13-JUN-1994;   94FR-0007191.
XX
PA   (INSR) INST ROUSSY GUSTAVE.
XX
PI   Bellet D, Chassin D, Koman A;
XX
DR   WPI; 1996-049682/05.
DR   N-PSDB; AAT10275.
XX
PT   Early placental insulin-like protein, EPIL/placentin - contains
PT   growth factor-like activity useful for e.g. promoting lactation or
PT   for regeneration of nerve, muscle, skin or bone tissue
XX
PS   Claim 1; Page 14-15; 25pp; French.
XX
CC   This is the amino acid sequence of a novel member of the insulin growth
CC   hormone family designated early placental insulin like (EPIL) protein or
CC   placentin. The encoding gene has been found to be expressed exclusively
CC   in the early placental tissue with a small amount in full term placental
CC   tissue. No other tissue expresses this gene. The gene was obtained from
CC   a cDNA library derived from mRNA isolated from cytotrophoblasts derived
CC   from first trimester placental tissue using, as a probe, a fragment of
CC   the gene amplified by primers AAT10276-7. Although the specific
CC   activity of the protein remains to be elucidated, it is thought that the
CC   protein will induce tyrosine phosphorylation of cellular proteins and may
CC   have growth factor-like activities e.g. human growth factor type 1 or
CC   lactation promoter activities. It may also be used to regenerate e.g.
CC   nerve, muscle, skin or bone tissue.
XX
SQ   Sequence   139 AA;

Alignment Scores:
Pred. No.:   1.19e-80
Score: 739.00   Length: 139
Percent Similarity: 100.00%   Matches: 139
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 95.48%   Mismatches: 0
RB: 17   Indels: 0
                                       Gaps: 0

US-09-518-842-1 (1-420) x AAR89134 (1-139)

Qy   1 ATGGCCAGCCTGTTCCGGTCCATCTGCCAGCAATCTGCTCGCTGCGCTGAGCGCAACTCCTT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b   1 MetAlaSerLeuPheArgSerTyrLeuProAlaIleTrpLeuLeuLeuSerGlnLeuLeu 20

Qy   61 AGAGAAAGCCTAGCAGCAGAGCTGAGGGGATGTGGTCCCGATTTGGAAAACACTTGCTG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b   21 ArgGluSerLeuAlaAlaGluLeuArgGlyCysGlyTyrProArgPheGlyLysHisLeuLeu 40

Qy   121 TCATATTGCCCATGCTGAGAAGACATTCACCACACCCCAGGAGGTGGCTGCTGGAA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b   41 SerTyrCysProMetProGluLysThrPheThrThrProGlyGlyTrpLeuLeuGlu 60

Qy   181 TCTGGACGTCCAAAGAAATGGTCTCAACCTCCAACAACAAAGATGGACAAGCCCTTAGGT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b   61 SerGlyArgProLysGluMetValSerThrSerAsnAsnLysAspGlyGlnAlaLeuGly 80
```

CC    in transformed host cells for use in the prepn. of antibodies and
CC    therapeutic compsns., or as a growth factor for maintaining cells
CC    in culture.  RRF-2 may be useful for modulating the reproductive
CC    physiology of mammals during pregnancy and parturition.
KX
SQ    Sequence    139 AA;

Alignment Scores:
Pred. No.:              1.19e-80          Length:           139
Score:                  739.00            Matches:          139
Percent Similarity:     100.00%           Conservative:     0
Best Local Similarity:  100.00%           Mismatches:       0
Query Match:            95.48%            Indels:           0
DB:                     18                Gaps:             0

US-09-518-842-1 (1-420) x AAW17676 (1-139)

Qy      1 ATGGCCAGCCTGTTCCGGTCCTATCTGCCAGCAATCTGGCTGCTGAGCCAACTCCTT  60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MetAlaSerLeuPheArgSerTyrLeuProAlaIleTrpLeuLeuSerGlnLeuLeu  20

Qy     61 AGAGAAAGCCTAGCAGCAGCAGCTGAGGGGATGGTGGTCCCCGATTTGGAAAACACTTGCTG  120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 ArgGluSerLeuAlaAlaAlaGluAlaArgGlyCysGlyProArgPheGlyLysHisHisLeu  40

Qy    121 TCATATTGCCCCATGCCTGAGAAGACATTCACCACACCCCAGAGAGGGTGGCTGCTGGAA  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     41 SerTyrCysProMetProGluLysThrPheThrThrProGlyGlyTrpLeuLeuGlu  60

Qy    181 TCTGGACGTCCCAAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGT  240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 SerGlyArgProLysGluMetValSerThrSerAsnAsnLysAspGlyGlnAlaLeuGly  80

Qy    241 ACGACATCAGAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGG  300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81 ThrThrSerGluPheIleProAsnLeuSerProGluLeuLysLysProLeuSerGluGly  100

Qy    301 CAGCCATCATTGAAGAAAATAAATACTTTCCCGCAAAAAGAGAAGTGGACGTCACAGATTT  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    101 GlnProSerLeuLysLysIleIleAsnLeuSerArgLysLysGluValAspValThrAsp  120

Qy    361 GATCCATTCTGTTGTGAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGTACA  417
          |||||||||||||||||||||||||||||||||||||||||||
Db    121 AspProPheCysGluValIleCysAspAspGlyThrSerValLysLeuCysThr  139

ESULT 3
AAW69168
ID    AAW69168 standard; Protein; 139 AA.
X
AC    AAW69168;
X
DT    07-OCT-1998  (first entry)
E     Zinsl protein.
X
W     Zinsl; human; placenta; placentin; pancreatic islet cell proliferation;
W     insulin secretion; diabetes; therapy.
X
S     Homo sapiens.
X
H     Key          Location/Qualifiers
T     Region       26..43
T                  /note= "B chain of Zinsl"
T     Region       115..0
T                  /note= "A chain of Zinsl"
X
N     WO9827210-A1.
X
D     25-JUN-1998.
X
R     16-DEC-1997;   97WO-US23326.
X
R     16-DEC-1996;   96US-0033003.

---

XX
PA    (ZYMO ) ZYMOGENETICS INC.
XX
PI    Conklin DC,  Hoffmann RC,   Humes JM,   Jaspers SR,   Ren HP;
PI    Sprugel KH;
XX
DR    WPI; 1998-362779/31.
DR    N-PSDB; AAV4463.
XX
PT    New isolated protein, Zinsl - obtained from human placenta, which
PT    increases the proliferation of pancreatic islet cells, used for
PT    treating diabetes
XX
PS    Claim 4; Page 60-61; 77pp; English.
XX
CC    This sequence is the human Zinsl protein of the invention. The Zinsl
CC    protein was isolated form human placenta, and is believed to be a new
CC    version of the mature protein of placentin, having disulphide bonded A
CC    and B chains. The protein can be used for stimulating the proliferation
CC    of pancreatic islets to increase insulin secretory capacity of mammals.
CC    In particular it can be used for the treatment of diabetes. It can also
CC    be used for stimulating in vitro proliferation of pancreatic islet cells.
CC    It can also be used for production of antibodies and in detection and
CC    diagnosis.
XX
SQ    Sequence    139 AA;

Alignment Scores:
Pred. No.:              1.19e-80          Length:           139
Score:                  739.00            Matches:          139
Percent Similarity:     100.00%           Conservative:     0
Best Local Similarity:  100.00%           Mismatches:       0
Query Match:            95.48%            Indels:           0
DB:                     19                Gaps:             0

US-09-518-842-1 (1-420) x AAW69168 (1-139)

Qy      1 ATGGCCAGCCTGTTCCGGTCCTATCTGCCAGCAATCTGGCTGCTGAGCCAACTCCTT  60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MetAlaSerLeuPheArgSerTyrLeuProAlaIleTrpLeuLeuSerGlnLeuLeu  20

Qy     61 AGAGAAAGCCTAGCAGCAGCAGCTGAGGGGATGGTGGTCCCCGATTTGGAAAACACTTGCTG  120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 ArgGluSerLeuAlaAlaAlaGluAlaArgGlyCysGlyProArgPheGlyLysHisHisLeu  40

Qy    121 TCATATTGCCCCATGCCTGAGAAGACATTCACCACACCCCAGAGAGGGTGGCTGCTGGAA  180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     41 SerTyrCysProMetProGluLysThrPheThrThrProGlyGlyTrpLeuLeuGlu  60

Qy    181 TCTGGACGTCCCAAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGT  240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 SerGlyArgProLysGluMetValSerThrSerAsnAsnLysAspGlyGlnAlaLeuGly  80

Qy    241 ACGACATCAGAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGG  300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81 ThrThrSerGluPheIleProAsnLeuSerProGluLeuLysLysProLeuSerGluGly  100

Qy    301 CAGCCATCATTGAAGAAAATAAATACTTTCCCGCAAAAAGAGAAGTGGACGATGGAACTTCAGTTAAATTATGTACA  360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    101 GlnProSerLeuLysLysIleIleLeuLeuSerArgLysLysGluProLeuLeuLysSerGluGluGly  120

Qy    361 GATCCATTCTGTTGTGAAGTAATTTGTGACGATGGAACTTCAGTAAATTATGTACA  417
          |||||||||||||||||||||||||||||||||||||||||||
Db    121 AspProPheCysGluValIleCysAspAspGlyThrSerValLysLeuCysThr  139

RESULT 4
AAY26926
ID    AAY26926 standard; Protein; 139 AA.
XX
AC    AAY26926;
XX
DT    21-DEC-1999  (first entry)
XX

DE   Human insulin-like 4 protein.
XX
KW   Human; insulin-like 4; INSL4; embryonic; c bone tissue; ligament;
KW   early-placental insulin-like protein; EPIL; antibody; probe; primer;
KW   diagnosis; pathology; differentiation; proliferation; cartilage;
KW   ossification; osteoporosis; dysplasia.
XX
OS   Homo sapiens.
XX
PN   WO9937780-A1.
XX
PD   29-JUL-1999.
XX
PF   22-JAN-1999;   99WO-FR00137.
XX
PR   23-JAN-1998;   98FR-0000715.
XX
PA   (INSR ) INST ROUSSY GUSTAVE.
XX
PI   Laurent A, Bellet D;
XX
DR   WPI; 1999-469135/39.
DR   N-PSDB; AAX83561.
XX
PT   Expression of the INSL4 gene in human embryonic bone tissue and
PT   ligaments, -
XX
PS   Disclosure; Fig 1; 60pp; French.
XX
CC   This sequence represents the human protein encoded by the insulin-like 4
CC   (INSL4) gene which is designated early-placental insulin-like (EPIL)
CC   protein. The INSL4 gene is expressed in human embryonic bone tissue and
CC   ligaments and encodes 3 different EPIL proteins designated EPIL 1, 2 or
CC   3. EPIL 1 is a single chain comprising amino acids 18-139, EPIL 2 is a
CC   2 chain protein with chain A comprising amino acids 115-139 and chain B
CC   comprising amino acids 18-58 and EPIL 3 is a single chain comprising
CC   amino acids 59-114. The nucleic acids and protein or antibodies against
CC   EPIL 1, 2 or 3, and probes or primers for INSL4 are useful for the
CC   diagnosis of pathology associated with abnormal differentiation and/or
CC   proliferation of bone tissue or ligaments or development of abnormal
CC   cartilage and/or abnormal ossification of forming bones, e.g.
CC   osteoporosis or dysplasia. Compositions capable of modulation
CC   differentiation, regeneration and/or proliferation of bone tissue
CC   and/or ligament cells are also useful in treating bone disease.
XX
SQ   Sequence   139 AA;

Alignment Scores:
Pred. No.:
Score:                 1.19e-80          Length:         139
Percent Similarity:    100.00%           Matches:        739
Best Local Similarity: 100.00%           Conservative:   0
Query Match:           95.48%            Mismatches:     0
B:                     20                Gaps:           0

US-09-518-842-1 (1-420) x AAY26926 (1-139)

y     1 ATGGCCAGCCTGTTCCGTCCATCGCTGCCAGCAATCGGCTGCTGCTGAGCCAACTCCTT 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b     1 MetAlaSerLeuPheArgSerTyrLeuProAlaIleTrpLeuLeuSerGlnLeuLeuLeu 20

y    61 AGAGAAGCCTAGCAGCAGCAGAGCTGAGGGGATGTGGTCCCCGATTGGAAAACACTTGCTG 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b    21 ArgGluSerLeuAlaAlaGluLeuArgLeuCysGlyProArgPheGlyLysHisLeuLeu 40

y   121 TCATATTGCCCCATGCCTGAGAGAGACATTCACCACCACCCCAGGAGGTGGCTGCTGGAA 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b    41 SerTyrCysProMetProGluLysThrPhePheThrThrThrProGlyGlyTrpLeuLeuGlu 60

y   181 TCTGGACGTCCCAAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTAGGT 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
b    61 SerGlyArgProLysGluMetValSerThrSerAsnAsnLysAspGlyGlnAlaLeuGly 80

Qy  241 ACGACATCAGAATTCATTCCTAATTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGG 300
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81 ThrThrSerGluPheIleProAsnLeuSerProGluLeuLysProLeuSerGluGluGly 100

Qy  301 CAGCCACCATTGAAGAAATAATATACTTCCGCAAAAGAGAAGGTGGACGTCACAGATTT 360
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  101 GlnProSerLeuLysLysIleIleLeuLeuSerArgLysArgSerGlyArgHisArgPhe 120

Qy  361 GATCCATTCTGTTGTGAAGTAATTGTGACGATGAAACTTCAGTAAATATGTACA 417
        |||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 AspProPheCysPheCysGluValIleCysAspAspGlyThrSerValLysLeuCysThr 139

RESULT 5
AAW99574
ID   AAW99574 standard; Protein; 139 AA.
XX
AC   AAW99574;
XX
DT   22-JUN-1999 (first entry)
XX
DE   Human early placental insulin-like (EPIL) polypeptide.
XX
KW   INSL-4; insulin-like gene; EPIL; early placental insulin-like; antibody;
KW   vector; probe; hybridisation; tumour; hypoglycaemia; hyperglycaemia;
KW   diabetes; cardiovascular.
XX
OS   Homo sapiens.
XX
PN   WO909172-A1.
XX
PD   25-FEB-1999.
XX
PF   12-AUG-1998;   98WO-FR01799.
XX
PR   03-NOV-1997;   97FR-0013802.
PR   14-AUG-1997;   97FR-0010387.
XX
PA   (INSR ) INST ROUSSY GUSTAVE.
XX
PI   Bellet D, Bidart JM, Troalen F, Mock P;
XX
DR   WPI; 1999-181038/15.
DR   N-PSDB; AAX27490.
XX
PT   EPIL polypeptides encoded by insulin-like gene 4  -  and corresponding
PT   nucleic acids, antibodies, probes, primers, etc.
XX
PS   Claim 1; Fig 1; 119pp; French.
XX
CC   This sequence represent the early placental insulin-like (EPIL) protein
CC   encoded by the INSL-4 (insulin-like gene 4) gene. The polypeptide,
CC   antibodies to the polypeptide, vectors containing the coding sequence and
CC   probes derived from the coding sequence, can be used to treat tumours,
CC   preferably angioproliferative tumours, especially Kaposi's sarcoma,
CC   tumours of the pancreas, liver, uterus or breast, angiosarcomas,
CC   glioblastomas, neuroblastomas, rhabdomyosarcomas or leiomyosarcomas,
CC   to promote vascularisation of specific tissues; to treat retinopathy,
CC   macular degeneration, psoriasis, endometriosis, rheumatoid arthritis,
CC   atherosclerosis or hyperthyroidism; to treat post-angioplastic
CC   restenosis; to promote or inhibit embryo implantation; to prevent and/or
CC   treat disorders directly or indirectly connected with insulin-like
CC   activity; to prevent and/or treat disorders directly or indirectly
CC   connected with a dysfunction in carbohydrate metabolism, especially
CC   connected with hypo glycaemia or hyperglycaemia, especially gestational
CC   diabetes and diabetic complications, especially cardiovascular
CC   complications.
XX
SQ   Sequence   139 AA;

Alignment Scores:
Pred. No.:
Score:                1.19e-80          Length:         139
Percent Similarity:   100.00%           Matches:        739.00
                                        Conservative:   0

CC          and B chains. The protein can be used for stimulating the proliferation
CC          of pancreatic islets to increase insulin secretory capacity of mammals.
CC          In particular it can be used for the treatment of diabetes. It can also
CC          be used for stimulating in vitro proliferation of pancreatic islet cells.
CC          It can also be used for production of antibodies and in detection and
CC          diagnosis.
XX
SQ          Sequence    124 AA;

Alignment Scores:
Pred. No.:                                          Length:            124
Score:                          3.14e-66            Matches:           114
Percent Similarity:             620.00              Conservative:      1
Best Local Similarity:          100.00%             Mismatches:        0
Query Match:                    99.13%              Indels:            0
DB:                             80.10%              Gaps:              0
                                19

US-09-518-842-1 (1-420)  x AAW69169 (1-124)

Qy      73 GCAGCAGAGCTGAGGGGATGGTGCTCCCGATTTGGAAAACACTTGCTGTGTCATATTGCCCC 132
           :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10 SerAlaGluLeuArgGlyCysGlyProArgPheGlyLysHisLeuLeuSerTyrCysPro  29

Qy     133 ATGCCTGAGAGACATTCACCACCACCCAGGAGGGTGCTGCTGGAATCTGGACGTCCC 192
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      30 MetProGluLysThrPheThrThrThrProGlyGlyTrpLeuGluSerGlyArgPro  49

Qy     193 AAAGAAATGGTGTCAACCTCCAACAACAAAGATGACAAGCCTTAGGTACGACATCAGAA 252
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50 LysGluMetValSerThrSerAsnAsnLysAspGlyGlnAlaLeuGlyThrThrSerGlu  69

Qy     253 TTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCATTG 312
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70 PheIleProAsnLeuSerProGluLeuLysLysProLeuSerGluGlyGlnProSerLeu  89

Qy     313 AAGAAAATAATACTTTCCCGCAAAAGAAGAAGTGACGTGACGATTTGATCCATTCTGT 372
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      90 LysLysIleIleLeuSerArgLysLysArgSerGlyArgHisArgPheAspPhePheCys 109

Qy     373 TGTGAAGTAATTTGTGACGATGGAACTTCAGTAAATTATGTACA 417
           ||||||||||||||||||||||||||||||||||||||||||||
Db     110 CysGluValIleCysAspAspGlyThrSerValLysLeuCysThr 124

RESULT 7
AAW69170
ID   AAW69170 standard; Protein; 159 AA.
XX
AC   AAW69170;
XX
DT   07-OCT-1998  (first entry)
XX
DE   N-terminally tagged Zinsl protein.
XX
KW   Zinsl; human; placenta; placentin; pancreatic islet cell proliferation;
KW   insulin secretion; diabetes; therapy.
XX
OS   Homo sapiens.
XX
PN   WO9827210-A1.
XX
PD   25-JUN-1998.
XX
PF   16-DEC-1997;   97WO-US23326.
XX
PR   16-DEC-1996;   96US-0033003.
XX
PA   (ZYMO ) ZYMOGENETICS INC.
XX
PI   Conklin DC, Hoffmann RC, Humes JM, Jaspers SR, Ren HP;
PI   Sprugel KH;
XX
DR   WPI; 1998-362779/31.
DR   N-PSDB; AAV44664.

---

Best Local Similarity:   100.00%           Mismatches:      0
Query Match:             95.48%            Indels:          0
DB:                                        Gaps:            0
                         20

US-09-518-842-1 (1-420)  x AAW99574 (1-139)

Qy       1 ATGGGCCAGGCTGTTCCGGTCCTATCTGCCAGCAATCTGCTGCTGAGCCAACTCCTT  60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MetAlaSerLeuPheArgSerTyrLeuProAlaIleTrpLeuLeuSerGlnLeuLeu  20

Qy      61 AGAGAAAGCCTAGCAGCAGAGCTGAGGGGATGGTGCTCCCGATTTGGAAAACACTTGCTG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21 ArgGluSerLeuAlaAlaGluLeuArgGlyCysGlyProArgPheGlyLysHisLeuLeu  40

Qy     121 TCATATTGCCCCATGCCTGAGAGACATTCACCACCACCCAGGAGGGTGGCTGCTGGAA 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41 SerTyrCysProMetProGluLysThrPheThrThrThrProGlyGlyTrpLeuLeuGlu  60

Qy     181 TCTGGACGTCCCAAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGT 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 SerGlyArgProLysGluMetValSerThrSerAsnAsnLysAspGlyGlnAlaLeuGly  80

Qy     241 ACGACATCAGAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGG 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81 ThrThrSerGluPheIleProAsnLeuSerProGluLeuLysLysProLeuSerGluGly 100

Qy     301 CAGCCATCATTGAAGAAAATAATACTTTCCCGCAAAAGAGAAGTGACGTGACAGATTT 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     101 GlnProSerLeuLysLysIleIleLeuSerArgLysLysArgSerGlyArgHisArgPhe 120

Qy     361 GATCCATTCTGTTGTGAAGTAATTTGTGACGATGGAACTTCAGTAAATTATGTACA 417
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 AspProPheCysCysGluValIleCysAspAspGlyThrSerValLysLeuCysThr 139

RESULT 6
AAW69169
ID   AAW69169 standard; Protein; 124 AA.
XX
AC   'AAW69169;
XX
DT   07-OCT-1998  (first entry)
XX
DE   Zinsl NF protein.
XX
KW   Zinsl; human; placenta; placentin; pancreatic islet cell proliferation;
KW   insulin secretion; diabetes; therapy; Zinsl NF.
XX
OS   Homo sapiens.
XX
PN   WO9827210-A1.
XX
PD   25-JUN-1998.
XX
PF   16-DEC-1997;   97WO-US23326.
XX
PR   16-DEC-1996;   96US-0033003.
XX
PA   (ZYMO ) ZYMOGENETICS INC.
XX
PI   Conklin DC, Hoffmann RC, Humes JM, Jaspers SR, Ren HP;
PI   Sprugel KH;
XX
CC   Example 3; Page 64; 77pp; English.
XX
CC   This sequence is the human Zinsl NF protein of the invention. The Zinsl
CC   protein was isolated form human placenta, and is believed to be a new
CC   version of the mature protein of placentin, having disulphide bonded A

---

New isolated protein, Zinsl - obtained from human placenta, which
increases the proliferation of pancreatic islet cells, used for
treating diabetes

XX
PT   New isolated protein, Zins1 - obtained from human placenta, which
PT   increases the proliferation of pancreatic islet cells, used for
PT   treating diabetes
XX
PS   Example 1; Page 62-63; 77pp; English.
XX
CC   This sequence is a N-terminally tagged version of the human Zins1
CC   protein of the invention. The Zins1 protein was isolated form human
CC   placenta, and is believed to be a new version of the mature protein of
CC   placenta, having disulphide bonded A and B chains. The protein can be
CC   used for stimulating the proliferation of pancreatic islets to increase
CC   insulin secretory capacity of mammals. In particular it can be used for
CC   the treatment of diabetes. It can also be used for stimulating in vitro
CC   proliferation of pancreatic islet cells. It can also be used for
CC   production of antibodies and in detection and diagnosis.
XX
SQ   Sequence  159 AA;

Alignment Scores:
Pred. No.:                                  Length:           159
Score:                    4.59e-66          Matches:          114
Percent Similarity:       619.00            Conservative:     0
Best Local Similarity:    100.00%           Mismatches:       0
Query Match:              100.00%           Indels:           0
                          79.97%            Gaps:             0
                          19

US-09-518-842-1 (1-420) x AAW69170 (1-159)

Qy    76 GCAGAGCTGAGGGGATGTGGTCCCCGATTTGAAAACACTTGCTGTCATATTGCCCCATG 135
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    46 AlaGluLeuArgGlyCysGlyTrpProArgPheGlyLysHisLeuLeuSerTyrCysProMet 65

Qy   136 CCTGAGAGACATTCACCACCACCAGGAGGGTGCTGCTGGAATCTGACGTCCCAAA 195
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    66 ProGluLysThrPhePheThrThrThrProGlyGlyTrpLeuLeuGluSerGlyArgProLys 85

Qy   196 GAAATGGTGTCAACCTCCAACAACAAAGATGACAAGCCTAGGTACGACGACATCAGAATTC 255
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    86 GluMetValSerThrSerAsnAsnLysAspGlyGlnAlaLeuGlyThrThrSerGluPhe 105

Qy   256 ATTCCTAATTTGTCACCAGAGCTGAAGAAACACTGTTCTGAAGGAGCCAGCCATCATTGAAG 315
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   106 IleProAsnLeuSerProGluLeuLysLysHisLeuSerGluGluGlyGlnProSerLeuLys 125

Qy   316 AAAATAATACTTTCCCGCAAAAAGAGAGAGTGGACGTCACAGATTTGATCCATTCTGTTGT 375
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   126 LysIleIleLeuSerArgLysLysArgSerGlyArgHisArgPheAspProPheCysCys 145

Qy   376 GAAGTAATTTGTGACGATGGAGACTTCAGTAAATTATGTACA 417
          |||||||||||||||||||||||||||||||||||||||
Db   146 GluValIleCysAspAspGlyThrSerValLysLeuCysThr 159

RESULT 8
AAP30392
D    AAP30392 standard; Protein; 182 AA.
X
C    AAP30392;
X
T    25-MAR-2003  (updated)
T    25-MAY-1992  (first entry)
X
E    Sequence of porcine preprorelaxin.
X
W    Relaxin; hormone.
X
S    Sus scrofa domestica.
X
N    EP86649-A.
X
D    24-AUG-1983.
X
F    11-FEB-1983;  83EP-0300714.

XX
PR   12-FEB-1982;    82AU-0002695.
PR   11-FEB-1983;    83AU-0011834.
XX
PA   (FLOR-) FLOREY INST EXP PHY.
PA   (FLOR-) FLOREY HOWARD INST.
XX
P-   Hudson PJ, Haley JD, Niall HD, Shine J;
XX
DR   WPI; 1983-748587/35.
DR   N-PSDB; AAN30196.
XX
PT   Genes and DNA transfer vectors for prorelaxin expression - useful
PT   in prodn. of porcine relaxin for veterinary and human use
XX
PS   Disclosure; Fig 5; 50pp; English.
XX
CC   The inventors claim synthetic porcine preprorelaxin and prorelaxin
CC   and synthetic A, B and C peptide chains of prolaxin, and a gene for
CC   expression of porcine preprorelaxin or prorelaxin, and their sub-
CC   units (see AAN30186). They also claim a double-stranded DNA fragment
CC   for the expression of the signal peptide chain of porcine
CC   preprorelaxin comprising a coding strand and a complementary strand
CC   corresp. to a defined mRNA sequence (see AAN30187-N30194) which
CC   corresp. to the most homologous regions between the pig and rat cDNA
CC   sequences. A probe (AAN30195) is also claimed.
CC   (Updated on 25-MAR-2003 to correct PP field.)
CC   (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ   Sequence  182 AA;

Alignment Scores:
Pred. No.:                                  Length:           182
Score:                    8.98e-13          Matches:          61
Percent Similarity:       180.00            Conservative:     12
Best Local Similarity:    38.02%            Mismatches:       55
Query Match:              31.77%            Indels:           64
                          23.26%            Gaps:             6
                          4

US-09-518-842-1 (1-420) x AAP30392 (1-182)

Qy     1 ATGGCCAGCCTGTTCCGGTCCCATCGTGCCAGCAATCTGGCTGCTGCTGAGCCAACTCCTT 60
          ||| |||||||| |||||    |||||||
Db     1 MetProArgLeuPhe---SerTyrLeuLeuGlyValTrpLeuLeuSerGlnLeuLeuPro 19

Qy    61 AGAGAA--------------------AGCCTAGCAGCAGAGCTGAGGGGGATGTGGTCCCCGATTTGGA 108
          ||||||                     :::::: |||
Db    20 ArgGluIleProGlyGlnSerThrAsnAspPheIleLysAlaCysGlyArgGluLeuVal 39

Qy   109 AAACACTTGCTGTCATATTGCCCCATGCCTGAGAAGACAATTCACCACCACCAGGAGGGG 168
          :::           ||||||            |||||  |||||
Db    91 ProAsnLeuProGlnGluLeuLysAlaThrLeuSerGluArgGlnProSerLeuArgGlu 110

Qy   169 TGG-----------------------------CTGCTGGAATCTGACGTCCCAAAGAA 198
          |||                             ||||||:::||  |||||||
Db    51 TrpGlyArgThrAlaLeuSerLeuGluGluProGlnLeuGluThrGlyProProAlaGlu 70

Qy   199 ATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACACATCAGAATTCATT 258
          |||||||||||        :::::| |||||||||
Db    71 ThrMetProSerSerIleThrLysAspAlaGluIleLeuLysMetLeuGluPheVal 90

Qy   259 CCTAATTTGTCACCAGAGCTGAAGAAACACTGTTCTGAAGGGCAGCCATCATTG------- 312
          ||||||||||||        |||||||
Db    91 ProAsnLeuProGlnGluLeuLysAlaThrLeuSerGluArgGlnProSerLeuArgGlu 110

Qy   313 ----------------------------------AAGAAAATAATA 324
          :::                                   |||   |||
Db   111 LeuGlnSerAlaSerLysAspSerAsnLeuAsnPheGluGluLysLysIleIleIle 130

Qy   325 CTT--------------------------------------------------- 327
          |||
Db   131 LeuAsnArgGlnAsnGluAlaGluAspLysSerLeuLeuLysAsnLeuGlyLeu 150

```
Qy    1 ATGGCCAGCCTGTTCCCGTCCTATCTGCCAGCAATCTGGCTGCTGCTGAGCAACTCCTT   60
        |||  |||  ||||||||||||||||        |:|||||||||||||||||||||||
Db    1 MetProArgLeuPhe---Ser***LeuLeuGlyValTrpLeuLeuLeuSerGlnLeuPro   19

Qy   61 AGAGAA-------------AGCCTAGCAGCAGAGCTGAGGGGATGGTGTCCCGATTGGA  108
        ||||||             :::::::
Db   20 ArgGluIleProGlyGlnSerThrAsnAspPheIleLysAlaCysGlyArgGluLeuVal   39

Qy  109 AAACACTTGCTGTCATATTGCCCCATGCCTGAGAAGACAATTCACCACCACCCAGGAGGG  168
        |||                :::                                  ------Ser  50
Db   40 ArgLeuTrpValGlyIleCysGlyGlySerVal------

Qy  169 TGG---------------------------CTGCTGGAATCTGGACGTCCCAAAGAA  198
        |||                                |||||:::|||
Db   51 TrpGlyArgThrAlaLeuSerGluGluGlnLeuProGlnLeuGluThrGlyProAlaGlu   70

Qy  199 ATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGTACGACATCAGAATTCATT  258
        ||          :: :::|||  ||  ::|||      |:||||:::
Db   71 ThrMetProSerSerIleThrLysAspAlaGluIleLeuLysMetMetLeuGluPheVal   90

Qy  259 CCTAATTGTCACCAGAGCTGAGAAAACCACTGTCTGAAGGGCAGCCATCATTG-----  312
        ||||||||||  |||||||||  ||||||||||||
Db   91 ProAsnLeuProGlnLeuGlyLeuGlyLysAlaThrLeuSerGluArgGlnProSerLeuArgGlu  110

Qy  313 ---------------------------------AAGAAAATAATA  324
                                              ||
Db  111 LeuGlnLeuGlnSerAlaSerLysAspSerAsnLeuAsnPheGluGluPheLysLysIleIle  130

Qy  325 CTT-----------------------------------------------  327
        |||
Db  131 LeuAsnArgGlnAsnGluAlaGluLysAspLysSerLeuLeuLeuLeuAsnLeuGlyLeu  150

Qy  328 -------TCCCGCAAAAGAGAAGTGGACGTCACAGATTTGATCCATTCTGTTGTGAA  378
               ||||||||||||||||  ||
Db  151 AspLysHisSerArgLysArgLysArgPheLeuArgSerThrLeuSerGluArgGluLysCysGln  170

Qy  379 GTAATTTGTGACGATGGAACTTCAGTTAAATTATGT  414
        ||  |||
Db  171 ValGlyCysIleArgLysAspIleAlaArgLeuCys  182
```

RESULT 9
AAP40156
ID   AAP40156 standard; Protein; 182 AA.
XX
AC   AAP40156;
XX
DT   25-MAR-2003 (updated)
DT   11-FEB-1992 (first entry)
XX
DE   Sequence of porcine preprorelaxin.
XX
GW   Labour; birth; hormone; relaxin.
XX
OS   Pig.
XX
FH   Key   Location/Qualifiers
FT   Peptide   1..25
FT   /label= signal
FT   Peptide   26..57
FT   /label= B-chain
FT   Peptide   58..160
FT   /label= C-peptide
FT   Peptide   161..182
FT   /label= A-chain
XX
FR   -EP101309-A.
XX
D   22-FEB-1984.
XX
FR   11-AUG-1983; 83EP-0304662.
XX
RR   12-AUG-1982; 82AU-0005352.
RR   11-AUG-1983; 83AU-0017906.
RR   01-JAN-1988; 88EP-0104503.
XX
A   (FLOR-) FLOREY HOWARD INST.
XX
I   Hudson PJ, Shine J, Niall HD, Tregear GW;
XX
RR   WPI; 1984-050918/09.
RR   N-PSDB; AAN40125, AAN40126.
XX
T   Genes for human relaxin, prorelaxin and preprorelaxin prodn. -
T   prepd. by recombinant DNA techniques
XX
S   Disclosure; Fig 3; 51pp; English.
XX
C   The inventors claim the gene for the expression of human
C   preprorelaxin (HPP) and its sub-units. Also claimed are synthetic
C   HPP, HP, fragments and analogues. Human relaxin and its analogues
C   are prepd. for therapeutic purposes, esp. in clinical intervention
C   in cases of difficult labour.
C   (Updated on 25-MAR-2003 to correct PA field.)
XX
Q   Sequence   182 AA.

Alignment Scores:
Pred. No.:   6.35e-12   Length: 182
Score: 173.00   Matches: 60
Percent Similarity: 37.50%   Conservative: 12
Best Local Similarity: 31.25%   Mismatches: 56
Query Match: 22.35%   Indels: 64
B: 5   Gaps: 6

S-09-518-842-1 (1-420) x AAP40156 (1-182)

RESULT 10
AAP94621
ID   AAP94621 standard; protein; 185 AA.
XX
AC   AAP94621;
XX
DT   25-MAR-2003 (updated)
DT   21-JUN-1990 (first entry)
XX
DE   Amino acid sequence of human preprorelaxin H2.
XX
KW   Relaxin; H2-relaxin; pubic symphysis.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   Peptide   1..25
FT   /label=Signal peptide
FT   Peptide   26..57
FT   /label=B-chain
FT   Peptide   58..161
FT   /label=C-peptide
FT   Peptide   162..159
FT   /label=A-chain
XX
PN   EP30033-A.
XX
PD   15-FEB-1989.
XX
PF   12-DEC-1983; 88EP-0110103.
XX
PR   13-DEC-1982; 82AU-0007247.

XX
PA (FLOR-) FLOREY INST EXPERIMENTAL PHYSIOLOGY.
XX
PI Hudson PJ, Tregear GW, Niall HD;
XX
DR WPI; 1989-047874/07.
DR N-PSDB; AAN92483.
XX
PT New human H2-relaxin analogues -
PT with shortened and/or modified A and/or B chains.
XX
PS Disclosure; Page -; 25pp; English.
XX
CC "H2" relaxin is deduced from a cDNA clone, has the general properties of
CC a growth factor and is capable of altering nature of connective tissue
CC and inducing smooth muscle contraction, specifically during labour.
CC (Updated on 25-MAR-2003 to correct PF field.)
CC (Updated on 25-MAR-2003 to correct PA field.)
CC (Updated on 25-MAR-2003 to correct PI field.)
XX
SQ Sequence 185 AA;

Alignment Scores:
Pred. No.:   6.48e-06   Length: 185
Score: 123.50   Matches: 53
Percent Similarity: 35.35%   Conservative: 17
Best Local Similarity: 26.77%   Mismatches: 55
Query Match: 15.96%   Indels: 73
DB: Gaps: 7

US-09-518-842-1 (1-420) x AAP94621 (1-185)

Qy   1 ATGGCCAGCCTGTTCCGGTCCTATCTGCCAGCAATCTGCTGCTGCTGAGCCAACTCCTT 60
            |||| | ||| ::::||
Db   1 MetProArgLeuPhePheHisLeuLeuGlyValCysLeuLeuAsnGlnPheSer 20

Qy   61 --AGAGAAAGCCCTA 72
                                                       :::
Db   21 ArgAlaValAlaAspSerTrpMetGluGluValCysGlyValIleLeuLysLeuCysGlyArgGluLeuVal 40

Qy   73 GCAGCAGAGCTGAGGGGATGGT---CCCGATTTGGAAAACACTTGCTGTCATATTGC 129
            |||||| |||||| ::: ||| |||||
Db   41 ArgAlaGlnIleAlaIleCysGlyMetSerThrTrpSerLysArgSerGlnSerGlnGlu 60

Qy 130 CCCATGCCTGAGAAGACCATTCACCACCCACCCGAGGAGGGTGGCGTGCTGCTGGAATCTGGACGT 189
            ||||||
Db   61 AspAlaProGln------------------------ThrPro----------Arg 67

Qy 190 CCC---AAAGAAATGGTTGTCAACCTCCAACAACAAGATGGACAAGCCTTAGGTACGACA 246
            ||:::||| ||||||| |||
Db   68 ProValAlaGluIleValProSerPheIleAsnLysAspThrGluThrIleAsnMetMet 87

Qy 247 TCAGAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCA 306
            |||:::|
Db   88 SerGluPheValAlaAsnLeuProGlnGluLeuLysLeuSerLeuThrIleuSerGluMetGlnPro 107

Qy 307 TCATTG-- 312
            :::||
Db 108 AlaLeuProGlnLeuGlnHisValProValLeuLysAspSerSerLeuLeuPheGlu 127

Qy 313 -----AAGAAAATATACTT-- 327
            ||:::||
Db 128 GluPheLysLysTyrLeuArgAsnArgGlnSerGluAlaAlaAspSerProSerGlu 147

Qy 328 -----------------------TCCCGCAAAAGAGAAGTGGACGTCACAGATTT 360
                                    ||||||||
Db 148 LeuLysTyrLeuGlyLeuAspThrHisSerArgLysLysAspSerLysLysAlaLeu 167

Qy 361 GATCCATTCTGTGTGAAGTAATTTGTGACGATGAACTTCAGTTAATTATGT 414
            |||||| || ::: |||
Db 168 AlaAsnLysCysCysHisValGlyCysThrLysTyrArgSerLeuAlaArgPheCys 185

AAP40108
ID AAP40108 standard; Protein; 185 AA.
XX
AC AAP40108;
XX
DT 25-MAR-2003 (updated)
DT 04-FEB-1992 (first entry)
XX
DE Sequence of human preprorelaxin H2.
XX
KW Relaxin; hormone; probe; uterine contraction.
XX
OS H2.
XX
FH Key   Location/Qualifiers
FT Peptide   1..24
FT   /label= signal
FT Region   25..57
FT   /label= B-chain
FT Region   58..161
FT   /label= C-peptide
FT Region   162..185
FT   /label= A-chain
XX
PN EP112149-A.
XX
PD 27-JUN-1984.
XX
PF 12-DEC-1983; 83EP-0307553.
XX
PR 13-DEC-1982; 82AU-0007247.
PR 09-DEC-1983; 83AU-0022283.
PR 01-JAN-1988; 88EP-0110103.
XX
PA (FLOR-) FLOREY INST EXPERIMENTAL PHYSIOLOGY.
XX
PI Hudson PJ, Niall HD, Tregear GW;
XX
DR WPI; 1984-160129/26.
DR N-PSDB; AAN40086.
XX
PT Gene for expression of human H2-preprorelaxin -
PT for prodn. of the peptide after expression etc.
XX
PS Disclosure; Fig 2; 56pp; English.
XX
CC The inventors claim a ds DNA SQ which comprises a coding strand and
CC a complementary strand corresp. to (AAN40086) and its subunits. Also
CC claimed are synthetic human H2-preprorelaxin, human H2-prorelaxin
CC and human H2-relaxin together with their analogues having shortened A
CC or B chains, or modified to contain different AA residues.
CC (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ Sequence 185 AA;

Alignment Scores:
Pred. No.:   1.5e-05   Length: 185
Score: 120.50   Matches: 49
Percent Similarity: 38.38%   Conservative: 22
Best Local Similarity: 26.49%   Mismatches: 67
Query Match: 15.57%   Indels: 47
DB: Gaps: 5

US-09-518-842-1 (1-420) x AAP40108 (1-185)

Qy   1 ATGGCCAGCCTGTTCCGGTCCTATCTGCCAGCAATCGGCTGCTGCTGAGCCAACTCCTT 60
            ||| | ||| :::::||
Db   1 MetProArgLeuPhePheHisLeuLeuGlyValCysLeuLeuAsnGlnPheSer 20

Qy   61 AGA--------GAAAGCCTAGCAGCAGAG---CTGAGGGGATGGTGTCCCCGATTTGGA 108
            ||| :::::|| ::::: |||
Db   21 ArgAlaValAlaAspSerTrpMetGluGluValCysLeuLysLeuCysGlyArgGluLeuVal 40

RESULT 11

```
XX
PT    Genes for human relaxin, prorelaxin and preprorelaxin prodn. -
PT    prepd. by recombinant DNA techniques
XX
PS    Disclosure; Pig 2; 51pp; English.
XX
CC    The inventors claim the gene for the expression of human
CC    preprorelaxin (HPP) and its sub-units. Also claimed are synthetic
CC    HPP, fragments and analogues. Human relaxin and its analogues
CC    are prepd. for therapeutic purposes, esp. in clinical intervention
CC    in cases of difficult labour.
CC    (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ    Sequence  185 AA;
```

```
Alignment Scores:
Pred. No.:              1.98e-05        Length:        185
Score:                  119.50          Matches:        52
Percent Similarity:     37.11%          Conservative:   20
Best Local Similarity:  26.80%          Mismatches:     57
Query Match:            15.44%          Indels:         65
DB:                                     Gaps:            7
```

```
US-09-518-842-1 (1-420) x AAP40154 (1-185)

Qy    1  ATGGCCAGCTGTCCGGTTCCTATCTGCCAGCAATCTGCTGTGCTGAGCCAACTCCTT   60
                                                          ..:|||
Db    1  MetProArgLeuPheLeuPheHisLeuLeuGluPheCysLeuLeuLeuAsnGlnPheSer  20

Qy   61  AGAGAAAGCCTAGCAGCAGAG----------CTGAGGGGATGTGGTCCCCGATTT      105
            |||   :::::|||:::      :::::  |||||
Db   21  Arg--AlaValAlaAlaLysTrpLysLeuLysGlyGlyArgGluLeu               39

Qy  106  GGAAAACACTTGCTGTCATATTGCCCATG----------------------CCTGAG    141
            :::       |||  |||                                :::
Db   40  ValArgAlaGlnIleAlaIleCysGlyMetSerThrTrpSerLysArgSerLeuSerGln  59

Qy  142  AAGACATTCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCC---AAAGAA    198
            ||||||||||                             :::  ||||||
Db   60  GluAspAlaProGlnAsnThrPro-------------------ArgProValAlaGlu    71

Qy  199  ATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCAGAATTCATT  258
            |||||      :::       |||||          :::          |||||
Db   72  IleValProSerPheIleAsnLysAspThrGluThrIleAsnMetSerGluPheIleLeu  91

Qy  259  CCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCATTG-----  312
            |||||  ||||||||||  |||||||||||||||
Db   92  AlaAsnLeuProGlnGlnLeuLysLeuAlaAlaLeuSerGluGluArgGlnProSerLeuProGlu  111

Qy  313  --------------------------------------------------AAGAAA    318
                                                              ||||||
Db  112  LeuGlnLeuTyrValProAlaLeuLysAspSerAsnLeuSerPheGluGluPheLysLys  131

Qy  319  ATAATACTTTCCCGC----------                                    333
            ::||||  :::||
Db  132  LeuIleArgAsnArgGlnSerGluAlaAlaAspSerAsnProSerGluLeuLysTyrLeu  151

Qy  334  --------------------------------------AAAAGAGAAGTGGACGTCACAGATTTGATCCATTCTGT  372
                                                  ||||||||
Db  152  GlyLeuAspThrHisSerGlnLysLysArgGlnLeuTyrSerArgAlaArgProTyrValAlaLeuPheGluGlyCys  171

Qy  373  TGTGAAGTAATTGTGACGATGGAACTTCAGTTAAATTATGT                   414
            :::      :::
Db  172  CysLeuIleGlyCysThrLysSerArgSerLeuAlaAlaTyrCys               185
```

```
Qy  109  AAACACTTGCTGTCATATTGCCCATGCTGAGAAGACATTCACCACCCCAGGAGGG    168
              :::   :::::|||  |||                 :::
Db   41  ArgAlaGlnIleAlaIleCysGlyMetSerThrTrpSerLysArgSerLeuSerGlnGlu  60

Qy  169  TGGCTGCTGGAATCTGGACGTCCC---AAAGAAATGTGTCAACCTCCAACAACAAAGAT   225
              :::::: |||||       :::  |||||
Db   61  AspAlaProGlnAsnProArgProValAlaGluIleValProSerPheIleAsnLysAsp  80

Qy  226  GGACAAGCCTTAGGTACGACATCAGAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAA  285
              ::: ||||||       |||||  :::  ||||||
Db   81  ThrGluThrIleAsnMetSerGluPheValAlaAlaAsnLeuProGlnGlnLeuLysLeu  100

Qy  286  CCACTGTCTGAAGGGCAGCCATCATTG--------------------------------  312
              |||||||  |||||::||
Db  101  ThrLeuSerGluMetGlnProAlaLeuProGlnLeuGlnGlnHisValProValLeuLys  120

Qy  313  --------------AAGAAAATAATACTT---------------------------    327
                          ||:|||||
Db  121  AspSerSerLeuPheGluGluPheLysLysLysLeuArgAsnArgGlnSerGluAla     140

Qy  328  ------------------------TCCCGCAAAAAG                          339
                                    ||||||||||
Db  141  AlaAspSerProSerGluLeuLysTyrLeuGlyLeuAspThrHisSerArgLysLys     160

Qy  340  AGAAGTGGACGTCACAGATTTGATCCATTCTGTGTGAAGTAATTGTGACGATGGAACT    399
              |||||||||  |||||||:::
Db  161  ArgGlnLeuTyrSerArgAlaLeuAlaLeuAlaAsnLysCysCysHisValGlyCysThrLysSer  180

Qy  400  TCAGTTAAATTATGT                                               414
              |||
Db  181  LeuAlaArgPheCys                                               185
```

```
RESULT 12
AAP40154
ID  AAP40154 standard; Protein; 185 AA.
XX
AC  AAP40154;
XX
DT  25-MAR-2003  (updated)
DT  11-FEB-1992  (first entry)
XX
DE  Sequence of human preprorelaxin.
XX
KW  Labour; birth; hormone; relaxin.
XX
OS  Homo sapiens.
XX
FH  Key            Location/Qualifiers
FT  Peptide        1..25
FT                 /label= signal
FT  Peptide        26..57
FT                 /label= B-chain
FT  Peptide        58..161
FT                 /label= C-peptide
FT  Peptide        162..185
FT                 /label= A-chain
XX
RN  EP101309-A.
RD  22-FEB-1984.
RN  11-AUG-1983;  83EP-0304662.
RN  12-AUG-1982;  82AU-0005352.
RN  11-AUG-1983;  83AU-0017906.
RN  01-JAN-1988;  88EP-0104503.
XX
PA  (FLOR-) FLORBY HOWARD INST.
XX
PI  Hudson PJ, Shine J, Niall HD, Tregear GW;
XX
PN  WPI; 1984-050918/09.
PN  N-PSDB; AAN40101.
```

```
RESULT 13
AAP40155
ID  AAP40155 standard; Protein; 185 AA.
XX
AC  AAP40155;
XX
DT  25-MAR-2003  (updated)
DT  11-FEB-1992  (first entry)
```

```
Qy   199 ATGGTGTCAACCTCCAACAACAAAGATGACAAGCCTTAGGTACGACATCAGAATTCATT 258
          ::::|||  ::: ||||||||||  ::   ::: |||||||||
Db    72 IleValProSerPheIleAsnLysAspThrGluThrIleIleIleMetLeuGluPheIle 91

Qy   259 CCTAATTTGTCACCAGAGCTGAAGAAACCACTGTTCTGAAGGGCAGCCATCATTG----- 312
          |||||| |||||||||||||||||||||||||||||||||||||||||||||
Db    92 AlaAsnLeuProGluLeuLysAlaAlaLeuSerGluArgGlnProSerLeuProGlu 111

Qy   313 ------------------------------------------------AAGAAA 318
                                                           |||||
Db   112 LeuGlnGlnTyrValProAlaLeuLysAspSerAsnLeuSerPheGluGluLysLys 131

Qy   319 ATAATACTTCCCGC-------------------------------------- 333
          ::::|| :::|||
Db   132 LeuIleArgAsnArgGlnSerGluAlaAlaAlaAspSerAsnProSerGluLeuLysTyrLeu 151

Qy   334 ------AAAAAGAGAAGTGACGTCACAGATTTGATCCATTCTGT 372
                |||||||||||||||||
Db   152 GlyLeuAspThrHisSerGlnLysLysArgArgProTyrValAlaLeuPheGluLysCys 171

Qy   373 TGTGAAGTAATTTGTGACGATGGAACTTCCAGTTAAAAATATGT 414
          |||  :::  |||
Db   172 CysLeuIleGlyCysThrLysArgSerLeuAlaLysTyrCys 185
```

```
RESULT 14
ABG61819
ID   ABG61819 standard; Protein; 185 AA.
XX
AC   ABG61819;
XX
DT   15-AUG-2002  (first entry)
XX
DE   Prostate cancer-associated protein #20.
XX
KW   Prostate cancer; prostate tumour tissue; human; mammal; cytostatic.
XX
OS   Mammalia.
XX
PN   WO200230268-A2.
XX
PD   18-APR-2002.
XX
PF   12-OCT-2001; 2001WO-US32045.
XX
PR   13-OCT-2000; 2000US-0687576.
PR   08-DEC-2000; 2000US-0733288.
PR   08-DEC-2000; 2000US-0733742.
PR   24-JAN-2001; 2001US-263957P.
PR   16-MAR-2001; 2001US-276791P.
PR   16-MAR-2001; 2001US-276888P.
PR   06-APR-2001; 2001US-281922P.
PR   24-APR-2001; 2001US-286214P.
PR   30-APR-2001; 2001US-0847046.
PR   04-MAY-2001; 2001US-288589P.
XX
PA   (EOSB-) EOS BIOTECHNOLOGY INC.
XX
PI   Gish KC, Mack DH, Wilson KE, Afar D, Hevezi P;
XX
DR   WPI; 2002-471335/50.
DR   N-PSDB; ABK92134.
XX
PT   Detecting a prostate cancer-associated transcript in a cell in a
PT   patient, useful for diagnosing prostate cancer (PC) or screening
PT   modulators of PC, by determining if prostate cancer-associated genes
PT   are expressed in a prostate tissue -
XX
PS   Claim 27; Page 315-316; 436pp; English.
XX
CC   The present invention relates to methods of detecting a prostate
CC   cancer-associated transcript in a cell from a patient. The method
CC   comprises contacting a biological sample from the patient with
```

```
XX
DE   Sequence of human preprorelaxin.
XX
CW   Labour; birth; hormone; relaxin.
XX
OS   Homo sapiens.
XX
KH   Key           Location/Qualifiers
FT   Peptide       1..25
FT                 /label= signal
FT   Peptide       26..57
FT                 /label= B-chain
FT   Peptide       58..161
FT                 /label= C-peptide
FT   Peptide       162..185
FT                 /label= A-chain
XX
PN   EP101309-A.
XX
PD   22-FEB-1984.
XX
PF   11-AUG-1983;  83EP-0304662.
XX
PR   12-AUG-1982;  82AU-0005352.
PR   11-AUG-1983;  83AU-0017906.
PR   01-JAN-1988;  88EP-0104503.
XX
PA   (FLOR-) FLOREY HOWARD INST.
XX
PI   Hudson PJ, Shine J, Niall HD, Tregear GW;
XX
DR   WPI; 1984-050918/09.
DR   N-PSDB; AAN40123, AAN40124.
XX
PT   Genes for human relaxin, prorelaxin and preprorelaxin prodn. -
PT   prepd. by recombinant DNA techniques
XX
PS   Disclosure; Fig 3; 51pp; English.
XX
CC   The inventors claim the gene for the expression of human
CC   preprorelaxin (HPP) and its sub-units. Also claimed are synthetic
CC   HPP, HP, fragments and analogues. Human relaxin and its analogues
CC   are prepd. for therapeutic purposes, esp. in clinical intervention
CC   in cases of difficult labour.
CC   (Updated on 25-MAR-2003 to correct PA field.)
XX
Q    Sequence  185 AA;
```

```
Alignment Scores:
Pred. No.:                            Length:        185
core:             119.50             Matches:        52
ercent Similarity: 37.11%            Conservative:   20
est Local Similarity: 26.80%         Mismatches:     57
uery Match:        15.44%            Indels:         65
B:                5                   Gaps:          7
```

```
S-09-518-842-1 (1-420) x AAP40155 (1-185)

Qy    1 ATGGCCAGCCTGTTCCGGTCCATACTTGCCAGCAATCTGCCTGCTGCTGAGCCAACTCCTT 60
         |||  ::: |||| ||||||  :::| |||||||||||||||  :::|||
Db    1 MetProArgLeuPheLeuPheHisLeuLeuPheCysLeuLeuLeuAsnGlnPheSer 20

Qy   61 AGAGAAAGCCTAGCAGCAGAG-----------------CTGAGGGGATGTGTCCCCGATTT 105
         |||  :::||:||||||||||                ::::|| |||||
Db   21 Arg--AlaValAlaAlaLysTrpLysAspAspValIleLysLeuCysGlyArgGluLeu 39

Qy  106 GGAAAACACTTGCTGTCATATTGCCCCATG----------------------CCTGAG 141
         ::|  ::::|||||| |||||||
Db   40 ValArgAlaGlyGlnIleAlaIleCysGlyMetSerThrTrpSerArgSerLeuSerGln 59

Qy  142 AAGACATTCACCACCACCCCAGGAGAGGGTGGCTGCTGGAATCTGGACGTCCC--AAAGAA 198
         :::  |||||| |||
Db   60 GluAspAlaProGlnThrPro------ArgProValAlaGlu 71
```

prostate cancer-associated polynucleotides (designated PC genes) that
selectively hybridise to a sequence that is at least 80% identical
to them. The prostate cancer-associated polynucleotide sequences
are differentially expressed in prostate tumour tissue or in
prostate cancer and are derived from the tissues of various
organisms such as humans or other mammals (e.g. mice, sheep and dogs).
The methods of the invention are useful for diagnosing and treating
prostate cancer in mammals. The prostate cancer-associated genes are
useful for diagnosing or treating prostate cancer, as well as for
identifying modulators of prostate cancer or agents that inhibit
prostate cancer. The nucleic acid sequences are particularly useful
in gene therapy, as a vaccine or in antisense applications.
ABG61800-ABG61944 represent prostate cancer-associated proteins.

Sequence   185 AA;

Alignment Scores:
Pred. No.:
Score: 1.98e-05
Percent Similarity: 119.50
Best Local Similarity: 37.11%
Query Match: 26.80%
DB: 15.44%
                          23

Length: 185
Matches: 52
Conservative: 20
Mismatches: 57
Indels: 65
Gaps: 7

US-09-518-842-1 (1-420) x ABG61819 (1-185)

```
Qy     1 ATGGCCAGCCTGTTCCGTCCTATCTGCCAGCAATCTGGCTGCTGCTGAGCCAACTCCTT  60
           |||   ||||||       ::||| | |||||||||||:::|||
Db     1 MetProArgLeuPheHisLeuGluPheCysLeuLeuLeuAsnGlnPheSer  20

Qy    61 AGAGAAAGCCTAGCAGCAGAG-----------------CTGAGGGGATGTGGTCCCCGATTT 105
           |||    :::||||||||||||:                ::::: |||||
Db    21 Arg--AlaValAlaAlaLysTrpLysAspValIleLysGlyArgGluLeu  39

Qy   106 GGAAAACACTTGCTGTCATATTGCCCCATG-------------CCTGAG 141
           ||||   ||| ||         |||   :::
Db    40 ValArgAlaGlnIleAlaIleCysGlyMetSerThrTrpSerLysSerGln  59

Qy   142 AAGACATTCACCACCACCCCCAGGAGGGTGGCTGCTGCTGGACGTCCC--AAAGA 198
           |||||                               ||||| |||
Db    60 GluAspAlaProGlnThrPro------------------ArgProValAlaGlu  71

Qy   199 ATGGTGTCAACCTCCAACAACAAGATGGACAAGCCTTAGGTACGACATCAGAATTCATT 258
           ::||| |       |||||||||   ::   ||||||||
Db    72 IleValProSerPheIleAsnLysAspThrGlnThrIleIleIleMetGluGluPheIle  91

Qy   259 CCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCATTG------ 312
           ||| |||||||||||  |||||||       ||||||||||||||
Db    92 AlaAsnLeuProProGluLeuLysLysAlaAlaLeuLeuLysGlnProSerLeuProGlu 111

Qy   313 ------------------------------------------AAGAAA 318
                                                       ||||||
Db   112 LeuGlnGlnTyrValProAlaLeuLysAspSerAsnLeuSerPheGluPheLysLys 131

Qy   319 ATAATACTTTCCCGC---------------------------------------------- 333
           |||||  :::|||
Db   132 LeuIleArgAspArgGlnSerGluAlaLeuAlaAspSerAsnProSerGluLeuLysTyrLeu 151

Qy   334 -----------------------AAAAAGAAGAAGTGGACGTCACAGATTTGATCCATTCTGT 372
                                   ||||||:::||||    |||
Db   152 GlyLeuAspThrHisSerGlnLysLysArgArgProTyrValAlaLeuPheGluLysCys 171

Qy   373 TGTGAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGT 414
           ||| |||   |||   :::    |||       ||
Db   172 CysLeuGlyCysThrLysSerLeuAlaLysTyrCys 185
```

SULT 15
LR64904
AAR64904 standard; Protein; 150 AA.

AAR64904;

25-MAR-2003   (updated)

DT   06-SEP-1995 (first entry)
XX
DE   Prorelaxin fragment amino acids 12-161.
XX
KW   Prorelaxin; cervical ripening; ovarian peptide hormone;
KW   mammary gland development; skin elasticity; cardiovascular therapy;
KW   relaxin; premature labour; plasmid pTR21.
XX
OS   Synthetic.
XX
XX
PN   WO9500645-A2.
XX
PD   05-JAN-1995.
XX
PF   20-JUN-1994; 94WO-US06997.
XX
PR   21-JUN-1993; 93US-0080354.
XX
PA   (GETH) GENENTECH INC.
XX
PI   Breece T, Hayenga K, Rinderknecht E,   Vandlen R;
PI   Yansura D;
XX
DR   WPI; 1995-052082/07.
DR   N-PSDB; AAQ76309.
XX
PT   Relaxin prodn. from non-natural recombinant prorelaxin - by
PT   cleavage of leader and C peptide chain, also new prorelaxin and
PT   related DNA, vectors etc.
XX
PS   Example 1; Fig 9A; 62pp; English.
XX
CC   AAQ76309 is the Not I-Bam HI restriction fragment of the plasmid
CC   pTR21, which encodes AAR64904, prorelaxin amino acids 12-161.
CC   Relaxin is produced by removing the non-naturally occuring leader
CC   and C-peptide from PR with a cleavage agent. Relaxin is an ovarian
CC   peptide hormone involved in the inhibition of premature labour,
CC   cervical ripening and the development of the mammary glands. It
CC   may also improve skin elasticity and has been used in cardiovascular
CC   therapy.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   150 AA;

Alignment Scores:
Pred. No.:
Score: 9.74e-05
Percent Similarity: 113.50
Best Local Similarity: 36.42%
Query Match: 27.78%
DB: 14.66%
                          16

Length: 150
Matches: 45
Conservative: 14
Mismatches: 46
Indels: 57
Gaps: 6

US-09-518-842-1 (1-420) x AAR64904 (1-150)

```
Qy    61 AGAGAAAGCCTAGCAGCAGAGCTGAGAGCTGAGGGGATGTGGT---CCCCGATTTGGAAAACACTTG 117
           ||||||   |||  |||||  ||||||:::||||
Db     2 ArgGluLeuValAlaArgAlaGlnIleAlaIleCysGlyMetSerThrTrpSerLysSerArgSer  21

Qy   118 CTGTCATATTGCCCCATGCCTGAAGACATTCACCACCACCCCCAGGAGGGTGGCTGCTGCTG 177
           ||||||:::       |||:::          |||||||
Db    22 LeuSerGlnGluAspAlaProGln------------------ThrPro------------  31

Qy   178 GAATCTGGACGTCCC---AAAGAAATGTGTCAACCTCCAACAACAAGATGGACAAGCC 234
           |||||||       |||||||:::     ||||||||||||||||
Db    32 -------ArgProValAlaGluIleValProSerPheIleAsnLysAspThrGluThr  48

Qy   235 TTAGGTACCGACATCAGAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCT 294
           :::  |||||||||:::   |||||||
Db    49 IleAsnMetMetSerGluPheValAlaAsnLeuProGluLeuLysLysLeuThrLeuSer  68

Qy   295 GAAGGGCAGCCATCATTG-------------------------------------- 312
           |||  |||   |||::::
Db    69 GluMetGlnProAlaLeuProGlnLeuGlnGlnHisValProValLeuLysAspSerSer  88
```

```
y   313 -----------------AAGAAATATACTT--------------- 327
                          |||||::::|||
b    89 LeuLeuPheGluGluPheLysLysLeuIleArgAsnArgGlnSerGluAlaAlaAspSer 108

y   328 --------------------------TCCCGCAAAAAGAGAAGTGGA 348
                                  ||||||||||||||||
b   109 SerProSerGluLeuLysTyrLeuGlyLeuAspThrHisSerArgLysLysArgGlnLeu 128

y   349 CGTCACAGATTTGATCCATTCTGTGTGAAGTAATTGTGACCATGGAACTTCAGTTAAA 408
             |||||| |||  |||            :::    :::
b   129 TyrSerAlaLeuAlaAsnLysCysCysHisValGlyCysThrLysArgSerLeuAlaArg 148

y   409 TTATGT 414
           |||
b   149 PheCys 150


earch completed: October  9, 2003, 12:15:11
ob time : 65.5 secs
```